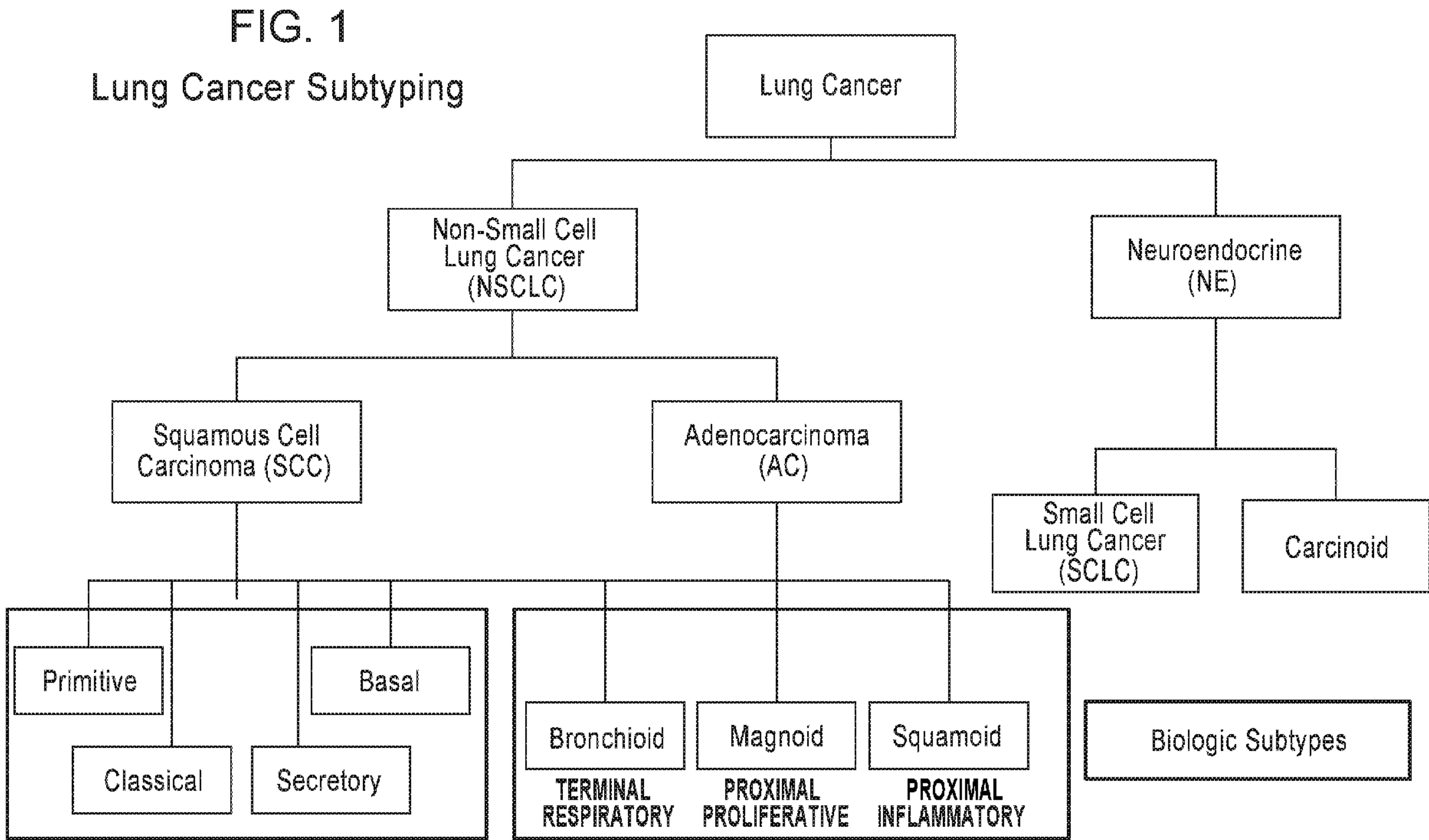




(12) **DEMANDE DE BREVET CANADIEN**  
**CANADIAN PATENT APPLICATION**  
(13) **A1**

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>(86) <b>Date de dépôt PCT/PCT Filing Date:</b> 2017/05/17</p> <p>(87) <b>Date publication PCT/PCT Publication Date:</b> 2017/11/23</p> <p>(85) <b>Entrée phase nationale/National Entry:</b> 2018/11/16</p> <p>(86) <b>N° demande PCT/PCT Application No.:</b> US 2017/033107</p> <p>(87) <b>N° publication PCT/PCT Publication No.:</b> 2017/201164</p> <p>(30) <b>Priorités/Priorities:</b> 2016/05/17 (US62/337,645);<br/>2016/08/25 (US62/379,402); 2016/09/19 (US62/396,587);<br/>2016/11/11 (US62/420,836); 2016/11/23 (US62/425,717)</p> | <p>(51) <b>Cl.Int./Int.Cl. C12Q 1/68</b> (2018.01),<br/><b>C07K 14/81</b> (2006.01), <b>C07K 16/30</b> (2006.01)</p> <p>(71) <b>Demandeurs/Applicants:</b><br/>GENECENTRIC THERAPEUTICS, INC., US;<br/>THE UNIVERSITY OF NORTH CAROLINA AT CHAPEL<br/>HILL, US</p> <p>(72) <b>Inventeurs/Inventors:</b><br/>FARUKI, HAWAZIN, US;<br/>LAI-GOLDMAN, MYLA, US;<br/>MAYHEW, GREG, US;<br/>SERODY, JONATHAN, US;<br/>PEROU, CHARLES, US;<br/>HAYES, DAVID NEIL, US</p> <p>(74) <b>Agent:</b> SMART &amp; BIGGAR</p> |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

(54) **Titre : PROCEDES DE SOUS-TYPAGE DU CARCINOME EPIDERMOIDE BRONCHIQUE**  
(54) **Title: METHODS FOR SUBTYPING OF LUNG SQUAMOUS CELL CARCINOMA**



(57) **Abrégé/Abstract:**  
Methods and compositions are provided for determining a subtype of lung squamous cell carcinoma (SQ) of an individual by detecting the expression level of at least one classifier biomarker selected from a group of gene signatures for lung squamous cell carcinoma. Also provided herein are methods and compositions for determining the response of an individual with a squamous cell carcinoma subtype to a therapy such as immunotherapy.

- (51) International Patent Classification:

C12Q 1/68 (2006.01) C07K 14/81 (2006.01)

C07K 16/30 (2006.01)
- (21) International Application Number:

PCT/US2017/033107
- (22) International Filing Date:

17 May 2017 (17.05.2017)
- (25) Filing Language:

English
- (26) Publication Language:

English
- (30) Priority Data:

62/337,645 17 May 2016 (17.05.2016) US

62/379,402 25 August 2016 (25.08.2016) US

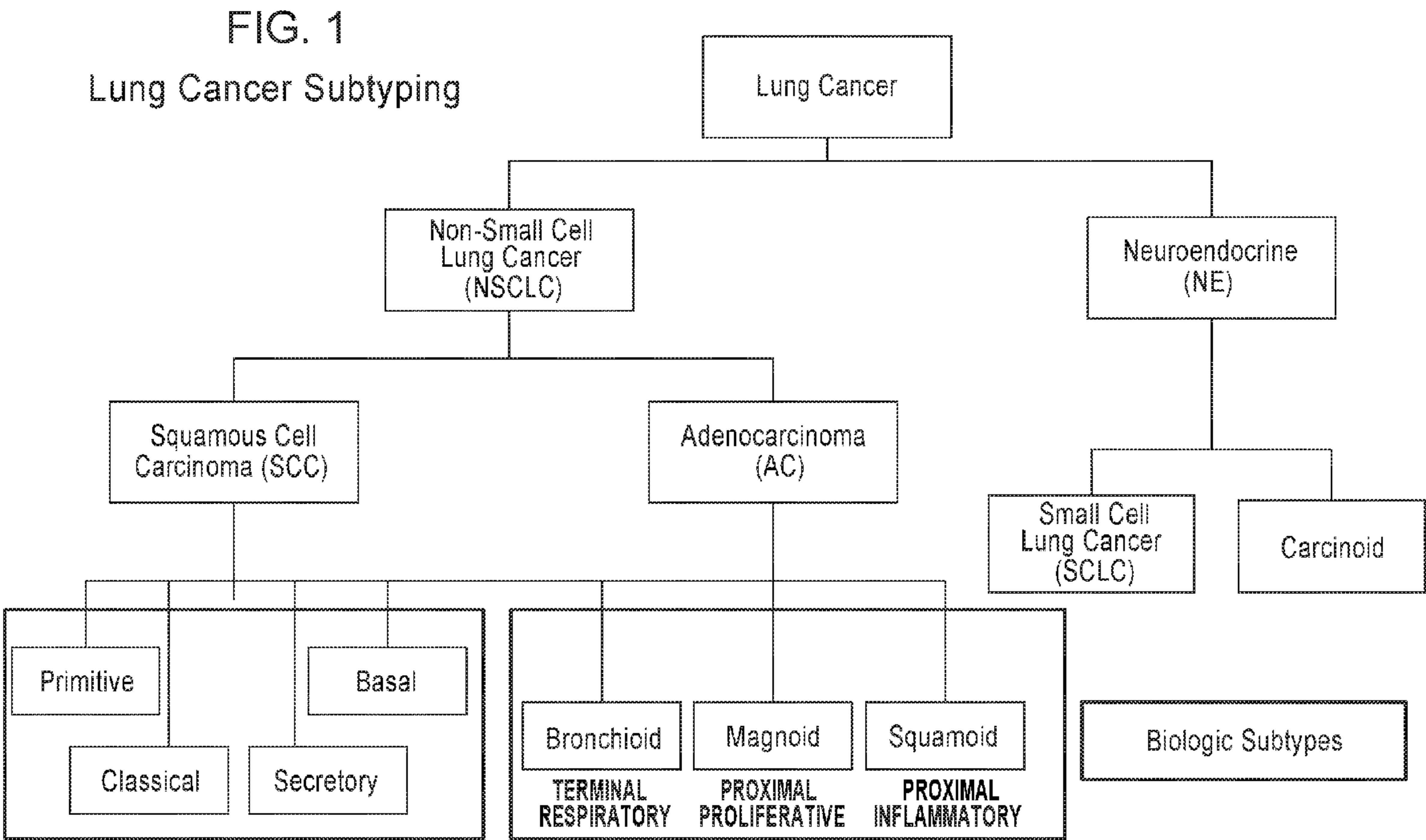
62/396,587 19 September 2016 (19.09.2016) US

62/420,836 11 November 2016 (11.11.2016) US

62/425,717 23 November 2016 (23.11.2016) US
- (71) Applicants: GENECENTRIC DIAGNOSTICS, INC.

[US/US]; 280 S. Mangum Street, Suite 350, Durham, North Carolina 27701 (US). THE UNIVERSITY OF NORTH
- CAROLINA AT CHAPEL HILL [US/US]; 100 Europa Drive, Suite 430, Chapel Hill, North Carolina 27517 (US).
- (72) Inventors: FARUKI, Hawazin; c/o GeneCentric Diagnostics, Inc., 280 S. Mangum Street, Suite 350, Durham, North Carolina 27701 (US). LAI-GOLDMAN, Myla; c/o GeneCentric Diagnostics, Inc., 280 S. Mangum Street, Suite 350, Durham, North Carolina 27701 (US). MAYHEW, Greg; c/o GeneCentric Diagnostics, Inc., 280 S. Mangum Street, Suite 350, Durham, North Carolina 27701 (US). SERODY, Jonathan; c/o GeneCentric Diagnostics, Inc., 280 S. Mangum Street, Suite 350, Durham, North Carolina 27701 (US). PEROU, Charles; c/o GeneCentric Diagnostics, Inc., 280 S. Mangum Street, Suite 350, Durham, North Carolina 27701 (US). HAYES, David Neil; c/o GeneCentric Diagnostics, Inc., 280 S. Mangum Street, Suite 350, Durham, North Carolina 27701 (US).
- (74) Agent: MCEVOY, Michael T. et al.; Cooley LLP, 1299 Pennsylvania Avenue, NW, Suite 700, Washington, District of Columbia 20004 (US).
- (81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM,

(54) Title: METHODS FOR SUBTYPING OF LUNG SQUAMOUS CELL CARCINOMA



(57) Abstract: Methods and compositions are provided for determining a subtype of lung squamous cell carcinoma (SQ) of an individual by detecting the expression level of at least one classifier biomarker selected from a group of gene signatures for lung squamous cell carcinoma. Also provided herein are methods and compositions for determining the response of an individual with a squamous cell carcinoma subtype to a therapy such as immunotherapy.

WO 2017/201164 A1 

---

AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

**(84) Designated States** *(unless otherwise indicated, for every kind of regional protection available):* ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

**Published:**

- *with international search report (Art. 21(3))*
- *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*
- *with sequence listing part of description (Rule 5.2(a))*

**METHODS FOR SUBTYPING OF LUNG SQUAMOUS CELL CARCINOMA****CROSS REFERENCE**

[0001] This application claims priority from U.S. Provisional Application No. 62/337,645 filed May 17, 2016, U.S. Provisional Application No. 62/379,402 filed August 25, 2016, U.S. Provisional Application No. 62/396,587 filed September 19, 2016, U.S. Provisional Application No. 62/420,836 filed November 11, 2016, and U.S. Provisional Application No. 62/425,717 filed November 23, 2016, each of which is incorporated by reference herein in its entirety for all purposes.

**FIELD OF THE INVENTION**

[0002] The present invention relates to methods for determining a squamous cell carcinoma subtype of a lung sample and for predicting the response to a treatment for a patient inflicted with specific types of lung cancer.

**STATEMENT REGARDING SEQUENCE LISTING**

[0003] The Sequence Listing associated with this application is provided in text format in lieu of a paper copy, and is hereby incorporated by reference into the specification. The name of the text file containing the Sequence Listing is GNCN\_010\_01WO\_SeqList\_ST25.txt. The text file is 319 KB, and was created on May 16, 2017, and is being submitted electronically via EFS-Web.

**BACKGROUND OF THE INVENTION**

[0004] Lung cancer is the leading cause of cancer deaths both in the United States and worldwide. Approximately 172,000 tumors of the lung were diagnosed in 2005 with an estimated 163,000 deaths, more than colon, breast, and prostate combined. At least 75% of patients present with locally advanced disease. Although there has been much effort to improve screening using technology such as high-resolution CT, these methods often produce false positive results and usually do not change outcome. Thus, even small tumors detected early present a significant threat to patients with postoperative 5-year survival rates for stage I lung cancer estimated between 47 to 63 percent. For patients with advanced disease the prognosis is worse with median survivals well under a year. In general, palliative therapy is effective but not sustainable and the average impact on overall survival is approximately 3 months.

[0005] At the population level the underlying cause of lung cancer is clearly tobacco use, with 90% of all lung cancers attributed directly to smoking. Smoking is so tightly correlated with lung cancer that it confounds definitive association with most other risk factors; although asbestos, radon, and a number of lung irritants are generally accepted as lung cancer risk factors. A genetic association is strongly suspected, however, the exact mechanism remains to be determined outside of a select group of rare Mendelian cancer syndromes. Despite many classification schemes and ongoing clinical trials, there has been overall disappointing progress in the field of clinical diagnostics and therapeutics.

[0006] Four distinct intrinsic lung squamous cell carcinoma subtypes exist that vary in their genomic profiles including gene expression, mutational spectrum, and copy number alterations. The four biologic lung squamous cell carcinoma subtypes, primitive, classical, secretory and basal, differ not only in their genomic features, but also demonstrate potentially important differences in clinical features.

[0007] Most lung cancers are classified as non-small cell lung carcinoma (NSCLC) (>85%), which is a diverse group with subtypes occurring throughout the respiratory tract. Adenocarcinoma (AD) and squamous cell carcinomas (SCC or SQ), the two main subtypes of NSCLC, are diagnosed at near equal frequency but are often found at different locations with SCC occurring more centrally. The 6th edition of the consensus classification of lung cancers developed by the World Health Organization (WHO) describes no fewer than 90 malignant morphologic classes and variants. There can often be heterogeneity, especially in larger tumors >1.5 cm, making morphological classification more difficult and leading to designations such as adeno-squamous carcinoma. Further, studies of histologic diagnosis reproducibility have shown limited intra-pathologist agreement and inter-pathologist agreement. Variability in morphology, limited tissue samples, and the need for assessment of a growing list of therapeutically targeted markers pose challenges to the current diagnostic standard. This is further highlighted by the idea that differentiation among various morphologic subtypes of lung cancer can be essential in guiding patient management and additional molecular testing can be used to identify specific therapeutic target markers.

[0008] Currently, gene expression based lung squamous cell carcinoma (SQ) subtyping has been primarily restricted to a research protocol involving the extraction of RNA from fresh frozen lung tumors, followed by application of a nearest centroid predictor using quantitative

gene expression of over 200 genes, as described in Wilkerson et al. Clin Cancer Res 2010; 16(19): 4864-75, which is herein incorporated by reference in its entirety.. Gene expression based SQ subtyping has been shown to classify squamous cell carcinoma tumors into 4 biologically distinct subtypes basal, classical, primitive and secretory. Further, these four subtypes can vary in their survival outcomes, patient populations, biological processes and in their immunogenic response features. Despite evidence of prognostic and predictive benefits from SQ subtyping, the requirement for gene expression of >200 genes in combination with complex bioinformatics analyses, has hindered the application of SQ subtyping in drug development and/or in the clinic.

[0009] Cancer immunosurveillance is the principle that the immune system can identify precancerous and cancerous cells and kill these cells before they become clinically relevant, which has been demonstrated in immunodeficient mouse models. Innate and adaptive immune responses can work together to either promote or inhibit cancer growth, and evasion of immune destruction is an emerging hallmark of cancer. Historically, methods of immune stimulation were not effective for lung cancer patients in the clinic. Deficiencies in tumor antigen expression and presentation on antigen presenting cells (APCs), infiltration of immunosuppressive cells and cytokines, and ineffective T-cell activation can lead to immunosuppression at the tumor site. Advances in the understanding of cancer and the immune system have led to effective therapies that activate antitumor responses, even in tumors that have highly developed methods of immune evasion, such as lung cancer. However the high immunosuppressive effects caused by lung tumors limit the beneficial effects of these advances due to a delicate balance between immunoactivation and immunosuppression in a patient. For example, in NSCLC, the role of immunosuppressive cells hampering immune activation is high, which is suggested to be related to the type of tumor, advanced stage of the disease, and the tumor load.

[0010] Therefore, developing a method to effectively distinguish intrinsic lung squamous cell carcinoma subtypes is critical for clinical diagnosis and disease management. Accordingly, new methods are needed to further define populations that might be likely to respond to immunotherapy. The present invention addresses these and other needs in the field for determining a prognosis or disease outcome for SQ patient populations based in part on the SQ subtype (basal, classical, primitive, secretory) of the patient. The methods of the invention provide a means for determining the cellular and molecular origins of lung cancer

(e.g., subtyping SQ) and can provide for more accurate diagnosis and applicable treatments as compared to diagnostic methods known in the art.

### **SUMMARY OF THE INVENTION**

[0011] In one aspect, provided herein is a method for determining a squamous cell carcinoma (SQ) subtype of a lung tissue sample obtained from a patient, the method comprising detecting an expression level of at least one classifier biomarker of **Table 1**, wherein the detection of the expression level of the classifier biomarker specifically identifies a basal, classical, secretory or primitive SQ subtype. In some cases, the method further comprises comparing the detected levels of expression of the at least one classifier biomarkers of **Table 1** to the expression of the at least one classifier biomarkers of **Table 1** in at least one sample training set(s), wherein the at least one sample training set comprises expression data of the at least one classifier biomarkers of **Table 1** from a reference SQ basal sample, expression data of the at least one classifier biomarkers of **Table 1** from a reference SQ classical sample, expression data of the at least one classifier biomarkers of **Table 1** from a reference SQ secretory sample, expression data of the at least one classifier biomarkers of **Table 1** from a reference SQ primitive sample or a combination thereof; and classifying the sample as basal, classical, secretory or primitive subtype based on the results of the comparing step. In some cases, the comparing step comprises applying a statistical algorithm which comprises determining a correlation between the expression data obtained from the sample and the expression data from the at least one training set(s); and classifying the sample as a basal, classical, secretory or primitive subtype based on the results of the statistical algorithm. In some cases, the expression level of the classifier biomarker is detected at the nucleic acid level. In some cases, the nucleic acid level is RNA or cDNA. In some cases, the detecting an expression level comprises performing quantitative real time reverse transcriptase polymerase chain reaction (qRT-PCR), RNAseq, microarrays, gene chips, nCounter Gene Expression Assay, Serial Analysis of Gene Expression (SAGE), Rapid Analysis of Gene Expression (RAGE), nuclease protection assays, Northern blotting, or any other equivalent gene expression detection techniques. In some cases, the expression level is detected by performing qRT-PCR. In some cases, the detection of the expression level comprises using at least one pair of oligonucleotide primers specific for at least one classifier biomarker of **Table 1**. In some cases, the sample is a formalin-fixed, paraffin-embedded (FFPE) lung tissue sample, fresh or a frozen tissue sample, an exosome, wash fluids, cell pellets, or a bodily fluid obtained from the patient. In some cases, the bodily fluid is blood or

fractions thereof, urine, saliva, or sputum. In some cases, the at least one classifier biomarker comprises a plurality of classifier biomarkers. In some cases, the plurality of classifier biomarkers comprises at least two classifier biomarkers, at least 10 classifier biomarkers, at least 20 classifier biomarkers, at least 30 classifier biomarkers, at least 40 classifier biomarkers, at least 50 classifier biomarkers, at least 60 classifier biomarkers, or at least 70 classifier biomarkers of **Table 1**. In some cases, the at least one classifier biomarker comprises all the classifier biomarkers of **Table 1**.

[0012] In another aspect, provided herein is a method for determining a squamous cell carcinoma (SQ) subtype of a lung tissue sample obtained from a patient comprising detecting an expression level of at least one nucleic acid molecule that encodes a classifier biomarker having a specific expression pattern in lung cancer cells, wherein the classifier biomarker is selected from the group consisting of the classifier genes set forth in **Table 1**, the method comprising: (a) isolating nucleic acid material from a lung tissue sample from a patient; (b) mixing the nucleic acid material with oligonucleotides that are substantially complementary to portions of nucleic acid molecule of the classifier biomarker; and (c) detecting expression of the classifier biomarker. In some cases, the method further comprises comparing the detected levels of expression of the at least one classifier biomarkers of **Table 1** to the expression of the at least one classifier biomarkers of **Table 1** in at least one sample training set(s), wherein the at least one sample training set comprises expression data of the at least one classifier biomarkers of **Table 1** from a reference SQ basal sample, expression data of the at least one classifier biomarkers of **Table 1** from a reference SQ classical sample, expression data of the at least one classifier biomarkers of **Table 1** from a reference SQ secretory sample, expression data of the at least one classifier biomarkers of **Table 1** from a reference SQ primitive sample or a combination thereof; and classifying the sample as basal, classical, secretory or primitive subtype based on the results of the comparing step. In some cases, the comparing step comprises applying a statistical algorithm which comprises determining a correlation between the expression data obtained from the sample and the expression data from the at least one training set(s); and classifying the sample as a basal, classical, secretory or primitive subtype based on the results of the statistical algorithm. In some cases, the detecting the expression level comprises performing qRT-PCR or any hybridization-based gene assays. In some cases, the expression level is detected by performing qRT-PCR. In some cases, the detection of the expression level comprises using at least one pair of oligonucleotide primers specific for at least one classifier biomarker of **Table 1**. In some cases, the method further comprises predicting the response to a therapy for treating a

subtype of lung squamous cell carcinoma (SQ) based on the detected expression level of the classifier biomarker. In some cases, the therapy is chemotherapy, angiogenesis inhibitors and/or immunotherapy. In some cases, the subtype of lung SQ is primitive and the therapy is an immunotherapy. In some cases, the sample is a formalin-fixed, paraffin-embedded (FFPE) lung tissue sample, fresh or a frozen tissue sample, an exosome, wash fluids, cell pellets or a bodily fluid obtained from the patient. In some cases, the bodily fluid is blood or fractions thereof, urine, saliva, or sputum. In some cases, the at least one nucleic acid molecule that encodes a classifier biomarker comprises a plurality of nucleic acid molecules that encode a plurality of classifier biomarkers. In some cases, the plurality of classifier biomarkers comprises at least two classifier biomarkers, at least 10 classifier biomarkers, at least 20 classifier biomarkers, at least 30 classifier biomarkers, at least 40 classifier biomarkers, at least 50 classifier biomarkers, at least 60 classifier biomarkers, or at least 70 classifier biomarkers selected from **Table 1**. In some cases, the at least one nucleic acid molecule that encodes a classifier biomarker comprises a plurality of nucleic acid molecules that include all the classifier biomarkers of **Table 1**.

**[0013]** In yet another aspect, provided herein is a method of detecting a biomarker in a lung tissue sample obtained from a patient, the method comprising measuring the expression level of a plurality of biomarker nucleic acids selected from **Table 1** using an amplification, hybridization and/or sequencing assay. In some cases, the lung tissue sample was previously diagnosed as being squamous cell carcinoma. In some cases, the previous diagnosis was by histological examination. In some cases, the amplification, hybridization and/or sequencing assay comprises performing quantitative real time reverse transcriptase polymerase chain reaction (qRT-PCR), RNAseq, microarrays, gene chips, nCounter Gene Expression Assay, Serial Analysis of Gene Expression (SAGE), Rapid Analysis of Gene Expression (RAGE), nuclease protection assays, Northern blotting, or any other equivalent gene expression detection techniques. In some cases, the expression level is detected by performing qRT-PCR. In some cases, the detection of the expression level comprises using at least one pair of oligonucleotide primers per each of the plurality of biomarker nucleic acids selected from **Table 1**. In some cases, the sample is a formalin-fixed, paraffin-embedded (FFPE) lung tissue sample, fresh or a frozen tissue sample, an exosome, wash fluids, cell pellets, or a bodily fluid obtained from the patient. In some cases, the bodily fluid is blood or fractions thereof, urine, saliva, or sputum. In some cases, the plurality of biomarker nucleic acids comprises, consists essentially of or consists of at least two biomarker nucleic acids, at least 10 biomarker nucleic acids, at least 20 biomarker nucleic acids, at least 30 biomarker nucleic

acids, at least 40 biomarker nucleic acids, at least 50 biomarker nucleic acids, at least 60 biomarker nucleic acids, or at least 70 biomarker nucleic acids of **Table 1**. In some cases, the plurality of biomarker nucleic acids comprises, consists essentially of or consists of all the classifier biomarker nucleic acids of **Table 1**.

**[0014]** In a further aspect, provided herein is a method of detecting a biomarker in a lung tissue sample obtained from a patient, the method consisting essentially of measuring the expression level of a plurality of biomarker nucleic acids selected from **Table 1** using an amplification, hybridization and/or sequencing assay. In some cases, the lung tissue sample was previously diagnosed as being squamous cell carcinoma. In some cases, the previous diagnosis was by histological examination. In some cases, the amplification, hybridization and/or sequencing assay comprises performing quantitative real time reverse transcriptase polymerase chain reaction (qRT-PCR), RNAseq, microarrays, gene chips, nCounter Gene Expression Assay, Serial Analysis of Gene Expression (SAGE), Rapid Analysis of Gene Expression (RAGE), nuclease protection assays, Northern blotting, or any other equivalent gene expression detection techniques. In some cases, the expression level is detected by performing qRT-PCR. In some cases, the detection of the expression level comprises using at least one pair of oligonucleotide primers per each of the plurality of biomarker nucleic acids selected from **Table 1**. In some cases, the sample is a formalin-fixed, paraffin-embedded (FFPE) lung tissue sample, fresh or a frozen tissue sample, an exosome, wash fluids, cell pellets, or a bodily fluid obtained from the patient. In some cases, the bodily fluid is blood or fractions thereof, urine, saliva, or sputum. In some cases, the plurality of biomarker nucleic acids comprises, consists essentially of or consists of at least two biomarker nucleic acids, at least 10 biomarker nucleic acids, at least 20 biomarker nucleic acids, at least 30 biomarker nucleic acids, at least 40 biomarker nucleic acids, at least 50 biomarker nucleic acids, at least 60 biomarker nucleic acids, or at least 70 biomarker nucleic acids of **Table 1**. In some cases, the plurality of biomarker nucleic acids comprises, consists essentially of or consists of all the classifier biomarker nucleic acids of **Table 1**.

**[0015]** In one aspect, provided herein is a method of detecting a biomarker in a lung tissue sample obtained from a patient, the method consisting of measuring the expression level of a plurality of biomarker nucleic acids selected from **Table 1** using an amplification, hybridization and/or sequencing assay. In some cases, the lung tissue sample was previously diagnosed as being squamous cell carcinoma. In some cases, the previous diagnosis was by histological examination. In some cases, the amplification, hybridization and/or sequencing assay comprises performing quantitative real time reverse transcriptase polymerase chain

reaction (qRT-PCR), RNAseq, microarrays, gene chips, nCounter Gene Expression Assay, Serial Analysis of Gene Expression (SAGE), Rapid Analysis of Gene Expression (RAGE), nuclease protection assays, Northern blotting, or any other equivalent gene expression detection techniques. In some cases, the expression level is detected by performing qRT-PCR. In some cases, the detection of the expression level comprises using at least one pair of oligonucleotide primers per each of the plurality of biomarker nucleic acids selected from **Table 1**. In some cases, the sample is a formalin-fixed, paraffin-embedded (FFPE) lung tissue sample, fresh or a frozen tissue sample, an exosome, wash fluids, cell pellets, or a bodily fluid obtained from the patient. In some cases, the bodily fluid is blood or fractions thereof, urine, saliva, or sputum. In some cases, the plurality of biomarker nucleic acids comprises, consists essentially of or consists of at least two biomarker nucleic acids, at least 10 biomarker nucleic acids, at least 20 biomarker nucleic acids, at least 30 biomarker nucleic acids, at least 40 biomarker nucleic acids, at least 50 biomarker nucleic acids, at least 60 biomarker nucleic acids, or at least 70 biomarker nucleic acids of **Table 1**. In some cases, the plurality of biomarker nucleic acids comprises, consists essentially of or consists of all the classifier biomarker nucleic acids of **Table 1**.

**[0016]** In another aspect, provided herein is a method of determining whether a squamous cell carcinoma patient is likely to respond to immunotherapy, the method comprising, determining the squamous cell carcinoma subtype of a lung tissue sample from the patient, wherein the squamous cell carcinoma subtype is selected from the group consisting of primitive, classical, secretory and basal; and based on the subtype, assessing whether the patient is likely to respond to immunotherapy. In some cases, the immunotherapy comprises checkpoint inhibitor therapy. In some cases, the checkpoint inhibitor targets PD-1 or PD-L1. In some cases, the checkpoint inhibitor targets CTLA-4. In some cases, the checkpoint inhibitor is Pembrolizumab, Nivolumab or an antigen fragment binding fragment thereof. In some cases, the checkpoint inhibitor is Ipilimumab or an antigen binding fragment thereof. In some cases, the patient is initially determined to have squamous cell carcinoma via a histological analysis of a sample. In some cases, the patient's squamous cell carcinoma subtype is selected from primitive, classical, secretory or basal, and is determined via a histological analysis of a sample obtained from the patient. In some cases, the sample is a formalin-fixed, paraffin-embedded (FFPE) lung tissue sample, fresh or a frozen tissue sample, an exosome, or a bodily fluid obtained from the patient. In some cases, the bodily fluid is blood or fractions thereof, urine, saliva, or sputum. In some cases, the determining the squamous cell carcinoma subtype comprises determining expression levels of a plurality of

classifier biomarkers. In some cases, the determining the expression levels of the plurality of classifier biomarkers is at a nucleic acid level by performing RNA sequencing, reverse transcriptase polymerase chain reaction (RT-PCR) or hybridization based analyses. In some cases, the plurality of classifier biomarkers for determining the squamous cell carcinoma subtype is selected from a publically available lung squamous cell carcinoma dataset. In some cases, the publically available lung squamous cell carcinoma dataset is TCGA Lung SQ RNAseq dataset. In some cases, the plurality of classifier biomarkers for determining the squamous cell carcinoma subtype is selected from **Table 1**. In some cases, the RT-PCR is quantitative real time reverse transcriptase polymerase chain reaction (qRT-PCR). In some cases, the RT-PCR is performed with primers specific to the plurality of classifier biomarkers of **Table 1**. In some cases, the method further comprises comparing the detected levels of expression of the plurality of classifier biomarkers of **Table 1** to the expression of the plurality of classifier biomarkers of **Table 1** in at least one sample training set(s), wherein the at least one sample training set comprises expression data of the plurality of classifier biomarkers of **Table 1** from a reference squamous cell carcinoma basal sample, expression data of the plurality of classifier biomarkers of **Table 1** from a reference squamous cell carcinoma classical sample, expression data of the plurality of classifier biomarkers of **Table 1** from a reference squamous cell carcinoma primitive sample, expression data of the plurality of classifier biomarkers of **Table 1** from a reference squamous cell carcinoma secretory sample or a combination thereof; and classifying the first sample as basal, classical, primitive or secretory based on the results of the comparing step. In some cases, the comparing step comprises applying a statistical algorithm which comprises determining a correlation between the expression data obtained from the sample and the expression data from the at least one training set(s); and classifying the sample as a basal, classical, primitive or secretory subtype based on the results of the statistical algorithm. In some cases, the plurality of the classifier biomarkers comprise each of the classifier biomarkers set forth in **Table 1**.

[0017] In yet another aspect, provided herein is a method for selecting a squamous cell carcinoma patient for immunotherapy, the method comprising, determining a squamous cell carcinoma subtype of a lung tissue sample from the patient, based on the subtype; and selecting the patient for immunotherapy. In some cases, the immunotherapy comprises checkpoint inhibitor therapy. In some cases, the checkpoint inhibitor targets PD-1 or PD-L1. In some cases, the checkpoint inhibitor targets CTLA-4. In some cases, the checkpoint inhibitor is Pembrolizumab, Nivolumab or an antigen fragment binding fragment thereof. In some cases, the checkpoint inhibitor is Ipilimumab or an antigen binding fragment thereof. In

some cases, the patient is initially determined to have squamous cell carcinoma via a histological analysis of a sample. In some cases, the patient's squamous cell carcinoma subtype is selected from primitive, classical, secretory or basal, and is determined via a histological analysis of a sample obtained from the patient. In some cases, the sample is a formalin-fixed, paraffin-embedded (FFPE) lung tissue sample, fresh or a frozen tissue sample, an exosome, or a bodily fluid obtained from the patient. In some cases, the bodily fluid is blood or fractions thereof, urine, saliva, or sputum. In some cases, the determining the squamous cell carcinoma subtype comprises determining expression levels of a plurality of classifier biomarkers. In some cases, the determining the expression levels of the plurality of classifier biomarkers is at a nucleic acid level by performing RNA sequencing, reverse transcriptase polymerase chain reaction (RT-PCR) or hybridization based analyses. In some cases, the plurality of classifier biomarkers for determining the squamous cell carcinoma subtype is selected from a publically available lung squamous cell carcinoma dataset. In some cases, the publically available lung squamous cell carcinoma dataset is TCGA Lung SQ RNAseq dataset. In some cases, the plurality of classifier biomarkers for determining the squamous cell carcinoma subtype is selected from **Table 1**. In some cases, the RT-PCR is quantitative real time reverse transcriptase polymerase chain reaction (qRT-PCR). In some cases, the RT-PCR is performed with primers specific to the plurality of classifier biomarkers of **Table 1**. In some cases, the method further comprises comparing the detected levels of expression of the plurality of classifier biomarkers of **Table 1** to the expression of the plurality of classifier biomarkers of **Table 1** in at least one sample training set(s), wherein the at least one sample training set comprises expression data of the plurality of classifier biomarkers of **Table 1** from a reference squamous cell carcinoma basal sample, expression data of the plurality of classifier biomarkers of **Table 1** from a reference squamous cell carcinoma classical sample, expression data of the plurality of classifier biomarkers of **Table 1** from a reference squamous cell carcinoma primitive sample, expression data of the plurality of classifier biomarkers of **Table 1** from a reference squamous cell carcinoma secretory sample or a combination thereof; and classifying the first sample as basal, classical, primitive or secretory based on the results of the comparing step. In some cases, the comparing step comprises applying a statistical algorithm which comprises determining a correlation between the expression data obtained from the sample and the expression data from the at least one training set(s); and classifying the sample as a basal, classical, primitive or secretory subtype based on the results of the statistical algorithm. In some cases, the plurality of the classifier biomarkers comprise each of the classifier biomarkers set forth in **Table 1**.

[0018] In one aspect, provided herein is a method of treating lung cancer in a subject, the method comprising: measuring the expression level of at least one biomarker nucleic acid in a lung cancer sample obtained from the subject, wherein the at least one biomarker nucleic acid is selected from a set of biomarkers listed in **Table 1**, wherein the presence, absence and/or level of the at least one biomarker indicates a subtype of the lung cancer; and administering an immunotherapeutic agent based on the subtype of the lung cancer. In some cases, the lung cancer sample is a squamous cell carcinoma sample, and wherein the set of biomarkers is **Table 1**. In some cases, the at least one biomarker nucleic acid selected from the set of biomarkers comprises, consists essentially of or consists of at least two biomarker nucleic acids, at least 10 biomarker nucleic acids, at least 20 biomarker nucleic acids, at least 30 biomarker nucleic acids, at least 40 biomarker nucleic acids, at least 50 biomarker nucleic acids, at least 60 biomarker nucleic acids, at least 70 biomarker nucleic acids or all of the biomarker nucleic acids of **Table 1**. In some cases, the lung tissue sample was previously diagnosed as being squamous cell carcinoma. In some cases, the previous diagnosis was by histological examination. In some cases, the method further comprises measuring the expression of at least one biomarker from an additional set of biomarkers. In some cases, the additional set of biomarkers comprise gene expression signatures of Innate Immune Cells (IIC), Adaptive Immune Cells (AIC), one or more individual immune biomarkers, one or more interferon(IFN) genes, one or more major histocompatibility complex, class II (MHCII) genes or a combination thereof. In some cases, the additional set of biomarkers comprises genes selected from **Tables 4A, 4B, 5, 6, 7**, or a combination thereof. In some cases, the gene expression signatures of AICs are selected from **Table 4A**. In some cases, the gene expression signature of IICs are selected from **Table 4B**. In some cases, the one or more individual immune biomarkers are selected from **Table 5**. In some cases, the one or more IFN genes are selected from **Table 6**. In some cases, the one or more MHCII genes are selected from **Table 7**. In some cases, the measuring the expression level is conducted using an amplification, hybridization and/or sequencing assay. In some cases, the amplification, hybridization and/or sequencing assay comprises performing quantitative real time reverse transcriptase polymerase chain reaction (qRT-PCR), RNAseq, microarrays, gene chips, nCounter Gene Expression Assay, Serial Analysis of Gene Expression (SAGE), Rapid Analysis of Gene Expression (RAGE), nuclease protection assays, Northern blotting, or any other equivalent gene expression detection techniques. In some cases, the expression level is detected by performing qRT-PCR. In some cases, the sample is a formalin-fixed, paraffin-embedded (FFPE) lung tissue sample, fresh or a frozen tissue sample, an exosome, wash

fluids, cell pellets, or a bodily fluid obtained from the patient. In some cases, the bodily fluid is blood or fractions thereof, urine, saliva, or sputum. In some cases, the subject's squamous cell carcinoma subtype is selected from primitive, classical, secretory or basal. In some cases, the lung cancer subtype is primitive and wherein the immunotherapeutic agent comprises a checkpoint inhibitor. In some cases, the checkpoint inhibitor targets PD-1 or PD-L1. In some cases, the checkpoint inhibitor targets CTLA-4. In some cases, the checkpoint inhibitor is Pembrolizumab, Nivolumab or an antigen fragment binding fragment thereof. In some cases, the checkpoint inhibitor is Ipilimumab or an antigen binding fragment thereof. In some cases, the at least one biomarker nucleic acid is a plurality of biomarker nucleic acids, wherein the plurality of biomarker nucleic acids comprises at least one biomarker nucleic acid listed in **Table 1** in combination with one or more biomarker nucleic acids from a publically available lung squamous cell carcinoma dataset, wherein the presence, absence and/or level of the plurality of biomarker nucleic acids indicates a subtype of the lung cancer. In some cases, the at least one biomarker nucleic acid is a plurality of biomarker nucleic acids, wherein the plurality of biomarker nucleic acids comprises all of the biomarker nucleic acids listed in **Table 1** in combination with one or more biomarker nucleic acids from a publically available lung squamous cell carcinoma dataset, wherein the presence, absence and/or level of the plurality of biomarker nucleic acids indicates a subtype of the lung cancer. In some cases, the publically available lung squamous cell carcinoma dataset is TCGA Lung SQ RNAseq dataset.

### **BRIEF DESCRIPTION OF THE FIGURES**

[0019] **FIG. 1** illustrates lung cancer subtyping and the biologic subtypes of squamous cell carcinoma (SCC or SQ) and Adenocarcinoma (AC or AD).

[0020] **FIG. 2** illustrates the lung SQ datasets used in the study described in Example 1.

[0021] **FIG. 3** illustrates a heatmap of immune cell signatures expression (i.e., Bindea et al reference from Example 1), other immune markers and individual immune markers in the Cancer Genome Atlas (TCGA) Lung SQ dataset.

[0022] **FIG. 4** illustrates correlation matrices of immune cell signatures in the TCGA SQ dataset where signatures were arranged by hierarchical clustering. Darker red means a pair of genes are more positively correlated. White means no correlation. Darker blue means negative correlation.

[0023] **FIG. 5** illustrates reproducibility of T cell signature gene expression subtype patterns across multiple SQ datasets as described in Example 1. RNAseq (Illumina, San Diego, CA) and microarrays from both Affymetrix (Santa Clara, CA) and Agilent (Santa Clara, CA).

[0024] **FIG. 6** illustrates association (adjusted R-squared) between CD274 (PD-L1) expression and adaptive immune cell (AIC) signatures in Squamous cell carcinoma (SCC or SQ) evaluation of the TCGA dataset. Association was consistently greater for subtypes than for PD-L1. In SQ, association was consistently greater for subtypes than for PD-L1 as described in Example 1. Tcm=central memory T cells, Tem=Effector Memory T cells, Th1=Type 1 T helper cells, Th2= Type 2 T helper cells, TFH=T follicular helper cells, Th17=T helper 17 cells, Treg= Tregulatory cells, Tgd=Gamma Delta Tcells.

[0025] **FIGs. 7A-7B** illustrate signature-survival associations overall and by subtype as described in Example 1. Hazard Ratios (HR) and confidence intervals calculated from stratified cox models correspond to a unit increase in the normalized immune marker and were adjusted for pathological stage. Subtype specific HR's were adjusted for stage (overall adjusted by stage and subtype) and only immune features with significant associations (nominal  $p < 0.05$ ) for at least one subtype are shown. SQ= Squamous Cell Carcinoma, MHC II= Major Histocompatibility Class II gene signature, Th1=Type 1 T helper cells, Th2= Type 2 T helper cells, TFH=T follicular helper cells, Th17= T helper 17 cells, Treg= Tregulatory cells, DC= Dendritic cells, iDC= Immature Dendritic Cells. **FIG. 7A-7B** is for SQ showing survival associations of immune cell signatures and markers by subtype in the TCGA cohort (**FIG. 7A**) or the TGCA, UNC and Raponi cohorts (**FIG. 7B**). Subtype specific immune marker hazard ratios and 95% confidence intervals were for 5 year overall survival in the TCGA cohort (n= 501 SQ) for **FIG. 7A**.

[0026] **FIG. 8** illustrates a comparison of approaches (i.e., 50-50 high/low approach on TCGA RNASeq lung SQ dataset; plain approach; hybrid approach; 50-50 high/low approach on transcriptome (50-50 high/low 20K approach)) for selecting genes for inclusion in a gene set for subtyping lung SQ as described in Example 2. A five-fold cross validation study was performed on performed on the Cancer Genome Atlas (TCGA) on a RNASeq lung squamous cell carcinoma (SQ) dataset except for the 50-50 high/low 20K approach

[0027] **FIG. 9** illustrates a five-fold cross validation study performed on the Cancer Genome Atlas (TCGA) RNASeq lung SQ dataset in order to determine an optimal number of genes to include for subtyping SQ.

[0028] **FIG. 10** illustrates gene-gene correlation coefficients and squamous cell carcinoma subtypes.

[0029] **FIG. 11** illustrates the gene rank (X-axis) and t-statistic (Y-axis) of genes from the 208 gene gold standard SQ classifier that were selected by application of a Classifying arrays to Nearest Centroid (CLaNC) algorithm with modification to the lung RNASeq SQ dataset (n = 506) from the Cancer Genome Atlas (TCGA) for inclusion in the SQ gene classifier set from **Table 1**.

[0030] **FIG. 12** illustrates the median gene expression of a subset of 20 genes from the 80 gene classifier selected for differentiating basal samples.

[0031] **FIG. 13** illustrates the median gene expression of a subset of 20 genes from the 80 gene classifier selected for differentiating classical samples.

[0032] **FIG. 14** illustrates the median gene expression of a subset of 20 genes from the 80 gene classifier selected for differentiating primitive samples.

[0033] **FIG. 15** illustrates the median gene expression of a subset of 20 genes from the 80 gene classifier selected for differentiating secretory samples.

[0034] **FIG. 16** illustrates agreement of SQ subtype prediction by the 80 gene signature (CLANC80) provided herein with the 208-gene classifier to define the gold standard subtype for multiple validation datasets and the newly collected FFPE validation dataset. The agreement with Lee, Raponi (rap), FFPE, TGCA and UNC is 84%, 91%, 87%, 88%, and 89%, respectively.

[0035] **FIG. 17** illustrates a heatmap of immune cell signatures (i.e., Bindea et al reference from Example 3), other immune markers and individual immune markers in the Cancer Genome Atlas (TCGA) Lung SQ dataset.

[0036] **FIG. 18** illustrates reproducibility of T cell signature gene expression subtype patterns across multiple SQ datasets as described in Example 3.

[0037] **FIG. 19** illustrates an association (adjusted R-squared) between CD274 (PD-L1) expression and adaptive immune cell (AIC) signatures versus subtype and AIC signatures as described in Example 3. Tem=central memory T cells, Tem=Effector Memory T cells, Th1=Type 1 T helper cells, Th2= Type 2 T helper cells, TFH=T follicular helper cells, Th17= T helper 17 cells, Treg= Tregulatory cells, Tgd=Gamma Delta Tcells.

[0038] **FIG. 20** illustrates for SQ signature-survival associations overall and by subtype as described in Example 3. Hazard Ratios (HR) and confidence intervals calculated from stratified cox models. Subtype specific HR's were adjusted for stage (overall adjusted by stage and subtype) and only immune features with significant associations ( $p < 0.05$ ) are shown.

[0039] **FIG. 21** illustrates box plots of all the immune cells and immunomarkers (i.e., IFN genes, MHCII genes and individual immunomarkers PDL1, PDL2, PDCD1 and CTLA4) by SQ subtype. SCC= squamous cell carcinoma.

[0040] **FIG. 22** illustrates Squamous cell carcinoma (SQ) subtype non-silent mutation burden, *NFE2L2* expression in SQ, and MHC class II signature, with Kruskal-Wallis association test p-values. MHC II= Major Histocompatibility Class II gene signature.

[0041] **FIG. 23** illustrates significant Squamous cell carcinoma (SQ) subtype differences in proliferation, non-silent mutation burden, and key drug targets: *CD274* (PD-L1), *PDCD1* (PD-1), and *CTLA4*. SQ subtyping was determined as described in Example 4.

[0042] **FIG. 24** illustrates significant drug target gene expression differences of SQ subtypes for the majority of genes in a clinical solid tumor mutation panel (322 genes disclosed in **Table 8**). In SQ subtypes, 76% showed differential expression (KW Bonferroni threshold  $p < 0.000155$ ). SQ subtyping was determined as described in Example 4.

[0043] **FIG. 25** illustrates significant drug target gene expression differences of SQ subtypes for the majority of genes in a clinical solid tumor mutation panel (322 genes disclosed in **Table 8**). In SQ subtypes, 70% showed differential expression (KW Bonferroni threshold  $p < 0.000155$ ). SQ subtyping was determined as described in Example 5.

[0044] **FIG. 26** illustrates significant Squamous cell carcinoma (SQ) subtype differences in proliferation. SQ subtyping was determined as described in Example 5.

## **DETAILED DESCRIPTION OF THE INVENTION**

### **Overview**

[0045] The present invention provides kits, compositions and methods for identifying or diagnosing lung cancer. That is, the methods can be useful for molecularly defining subsets of lung cancer, specifically lung squamous cell carcinoma (SQ). The methods provide a classification of lung cancer that can be prognostic and predictive for therapeutic response. While a useful term for epidemiologic purposes, "lung cancer" may not refer to a specific disease, but rather can represent a heterogeneous collection of tumors of the lung, bronchus, and pleura. For practical purposes, lung cancer can generally be divided into two histological subtypes-small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC). These main tumor types can present at different frequencies, can have different anatomic locations, can have different predilections for metastasis, may respond differently to therapy, and may likely be derived from different cell progenitors.

[0046] "Determining a squamous cell carcinoma subtype" can include, for example, diagnosing or detecting the presence and type of lung squamous cell carcinoma, monitoring the progression of the disease, and identifying or detecting cells or samples that are indicative of subtypes.

[0047] In one embodiment, lung cancer status is assessed through the evaluation of expression patterns, or profiles, of a plurality of classifier genes or biomarkers in one or more subject samples. For the purpose of discussion, the term "subject", or "subject sample", refers to an individual regardless of health and/or disease status. A subject can be a subject, a study participant, a patient, a control subject, a screening subject, or any other class of individual from whom a sample is obtained and assessed in the context of the invention. Accordingly, a subject can be diagnosed with lung squamous cell carcinoma (including subtypes, or grades thereof), can present with one or more symptoms of lung SQ cancer, or a predisposing factor, such as a family (genetic) or medical history (medical) factor, for lung cancer, can be undergoing treatment or therapy for lung cancer, or the like. Alternatively, a subject can be healthy with respect to any of the aforementioned factors or criteria. It will be appreciated that the term "healthy" as used herein, is relative to lung cancer status, as the term "healthy" cannot be defined to correspond to any absolute evaluation or status. Thus, an individual defined as healthy

with reference to any specified disease or disease criterion, can in fact be diagnosed with any other one or more diseases, or exhibit any other one or more disease criterion, including one or more other cancers.

**[0048]** As used herein, an “expression profile” or a “biomarker profile” or “gene signature” comprises one or more values corresponding to a measurement of the relative abundance, level, presence, or absence of expression of a biomarker or a discriminative or classifier gene. An expression profile can be derived from a subject prior to or subsequent to a diagnosis of lung cancer, can be derived from a biological sample collected from a subject at one or more time points prior to or following treatment or therapy, can be derived from a biological sample collected from a subject at one or more time points during which there is no treatment or therapy (e.g., to monitor progression of disease or to assess development of disease in a subject diagnosed with or at risk for lung cancer), or can be collected from a healthy subject. The term subject can be used interchangeably with patient. The patient can be a human patient. The one or more biomarkers of the biomarker profiles provided herein are selected from one or more biomarkers of **Table 1**.

**[0049]** As used herein, the term "determining an expression level" or "determining an expression profile" or “detecting an expression level” or “detecting an expression profile” as used in reference to a biomarker or classifier means the application of a biomarker specific reagent such as a probe, primer or antibody and/or a method to a sample, for example a sample of the subject or patient and/or a control sample, for ascertaining or measuring quantitatively, semi-quantitatively or qualitatively the amount of a biomarker or biomarkers, for example the amount of biomarker polypeptide or mRNA (or cDNA derived therefrom). For example, a level of a biomarker can be determined by a number of methods including for example immunoassays including for example immunohistochemistry, ELISA, Western blot, immunoprecipitation and the like, where a biomarker detection agent such as an antibody for example, a labeled antibody, specifically binds the biomarker and permits for example relative or absolute ascertaining of the amount of polypeptide biomarker, hybridization and PCR protocols where a probe or primer or primer set are used to ascertain the amount of nucleic acid biomarker, including for example probe based and amplification based methods including for example microarray analysis, RT-PCR such as quantitative RT-PCR (qRT-PCR), serial analysis of gene expression (SAGE), Northern Blot, digital molecular barcoding technology, for example Nanostring Counter Analysis, and TaqMan quantitative PCR assays.

Other methods of mRNA detection and quantification can be applied, such as mRNA in situ hybridization in formalin-fixed, paraffin-embedded (FFPE) tissue samples or cells. This technology is currently offered by the QuantiGene ViewRNA (Affymetrix), which uses probe sets for each mRNA that bind specifically to an amplification system to amplify the hybridization signals; these amplified signals can be visualized using a standard fluorescence microscope or imaging system. This system for example can detect and measure transcript levels in heterogeneous samples; for example, if a sample has normal and tumor cells present in the same tissue section. As mentioned, TaqMan probe-based gene expression analysis (PCR-based) can also be used for measuring gene expression levels in tissue samples, and this technology has been shown to be useful for measuring mRNA levels in FFPE samples. In brief, TaqMan probe-based assays utilize a probe that hybridizes specifically to the mRNA target. This probe contains a quencher dye and a reporter dye (fluorescent molecule) attached to each end, and fluorescence is emitted only when specific hybridization to the mRNA target occurs. During the amplification step, the exonuclease activity of the polymerase enzyme causes the quencher and the reporter dyes to be detached from the probe, and fluorescence emission can occur. This fluorescence emission is recorded and signals are measured by a detection system; these signal intensities are used to calculate the abundance of a given transcript (gene expression) in a sample.

[0050] In one embodiment, the “expression profile” or a “biomarker profile” or “gene signature” associated with the gene cassettes or classifier genes described herein (e.g., **Tables 1** and **2**) can be useful for distinguishing between normal and tumor samples. In another embodiment, the tumor samples are lung SQ samples. In another embodiment, SQ can be further classified as basal, classical, primitive or secretory based upon an expression profile determined using the methods provided herein. The characterization of basal, classical, primitive or secretory squamous cell carcinoma using gene expression has been described in Wilkerson MD, et al. Clin Cancer Res 2010; 16(19): 4864-75.

[0051] Expression profiles using the classifier or biomarker genes disclosed herein (e.g., **Table 1**) can provide valuable molecular tools for specifically identifying lung squamous cell carcinoma subtypes, and for evaluating therapeutic efficacy in treating lung squamous cell carcinoma. Accordingly, the invention provides methods for screening and classifying a subject for molecular SQ subtypes and methods for monitoring efficacy of certain therapeutic treatments for lung SQ.

[0052] In some instances, a single classifier gene provided herein is capable of identifying subtypes of lung squamous cell carcinoma with a predictive success of at least about 70%, at least about 71%, at least about 72%, about 73%, about 74%, about 75%, about 76%, about 77%, about 78%, about 79%, about 80%, about 81%, about 82%, about 83%, about 84%, about 85%, about 86%, about 87%, about 88%, about 89%, about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, about 99%, up to 100%.

[0053] In some instances, a single classifier gene as provided herein is capable of determining lung squamous cell carcinoma subtypes with a sensitivity or specificity of at least about 70%, at least about 71%, at least about 72%, about 73%, about 74%, about 75%, about 76%, about 77%, about 78%, about 79%, about 80%, about 81%, about 82%, about 83%, about 84%, about 85%, about 86%, about 87%, about 88%, about 89%, about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, about 99%, up to 100%.

[0054] The present invention also encompasses a system capable of distinguishing various subtypes of lung squamous cell carcinoma not detectable using current methods. This system can be capable of processing a large number of subjects and subject variables such as expression profiles and other diagnostic criteria. The methods described herein can also be used for "pharmacometabonomics," in analogy to pharmacogenomics, e.g., predictive of response to therapy. In this embodiment, subjects could be divided into "responders" and "nonresponders" using the expression profile as evidence of "response," and features of the expression profile could then be used to target future subjects who would likely respond to a particular therapeutic course.

[0055] The expression profile can be used in combination with other diagnostic methods including histochemical, immunohistochemical, cytologic, immunocytologic, and visual diagnostic methods including histologic or morphometric evaluation of lung tissue.

[0056] In various embodiments of the present invention, the expression profile derived from a subject is compared to a reference expression profile. A "reference expression profile" can be a profile derived from the subject prior to treatment or therapy; can be a profile produced from the subject sample at a particular time point (usually prior to or following treatment or therapy, but can also include a particular time point prior to or following

diagnosis of lung cancer); or can be derived from a healthy individual or a pooled reference from healthy individuals. A reference expression profile can be generic for lung cancer, or can be specific to different subtypes of lung squamous cell carcinoma.

[0057] The reference expression profile can be compared to a test expression profile. A "test expression profile" can be derived from the same subject as the reference expression profile except at a subsequent time point (e.g., one or more days, weeks or months following collection of the reference expression profile) or can be derived from a different subject. In summary, any test expression profile of a subject can be compared to a previously collected profile from a subject that has a basal, classical, primitive or secretory subtype.

[0058] The classifier biomarkers of the invention can include nucleic acids (RNA, cDNA, and DNA) and proteins, and variants and fragments thereof. Such biomarkers can include DNA comprising the entire or partial sequence of the nucleic acid sequence encoding the biomarker, or the complement of such a sequence. The biomarkers described herein can include RNA comprising the entire or partial sequence of any of the nucleic acid sequences of interest, or their non-natural cDNA products, obtained synthetically *in vitro* in a reverse transcription reaction. The biomarker nucleic acids can also include any expression product or portion thereof of the nucleic acid sequences of interest. A biomarker protein can be a protein encoded by or corresponding to a DNA biomarker of the invention. A biomarker protein can comprise the entire or partial amino acid sequence of any of the biomarker proteins or polypeptides. The biomarker nucleic acid can be extracted from a cell or can be cell free or extracted from an extracellular vesicular entity such as an exosome.

[0059] A "classifier biomarker" or "biomarker" or "classifier gene" can be any gene or protein whose level of expression in a tissue or cell is altered compared to that of a normal or healthy cell or tissue. For example, a "classifier biomarker" or "biomarker" or "classifier gene" can be any gene or protein whose level of expression in a tissue or cell is altered in a specific lung squamous cell carcinoma subtype. The detection of the biomarkers of the invention can permit the determination of the specific subtype. The "classifier biomarker" or "biomarker" or "classifier gene" may be one that is up-regulated (e.g. expression is increased) or down-regulated (e.g. expression is decreased) relative to a reference or control as provided herein. The reference or control can be any reference or control as

provided herein. In some embodiments, the expression values of genes that are up-regulated or down-regulated in a particular subtype of lung squamous cell carcinoma can be pooled into one gene cassette. The overall expression level in each gene cassette is referred to herein as the "expression profile" and is used to classify a test sample according to the subtype of lung squamous cell carcinoma. However, it is understood that independent evaluation of expression for each of the genes disclosed herein can be used to classify tumor subtypes without the need to group up-regulated and down-regulated genes into one or more gene cassettes. In some cases, as shown in **Table 2**, a total of 80 biomarkers can be used for SQ subtype determination. For each SQ subtype, 10 of the 20 biomarkers can be negatively correlated genes while 10 can be positively correlated genes which can be selected as the gene signature of a specific SQ subtype.

[0060] The classifier biomarkers of the invention can include any gene or protein that is selectively expressed in lung SQ, as defined herein above. Sample biomarker genes are listed in **Table 1** or **2**, below. In **Table 2**, the first column of the table represents the biomarker list selected for distinguishing basal SQ. The second column of the table represents the biomarker list selected for classical SQ. The third column of the table represents the biomarker list selected for distinguishing primitive SQ. The fourth column of the table represents the biomarker list selected for distinguishing secretory SQ.

[0061] The relative gene expression levels as represented by the tsat as described herein of the classifier biomarkers for lung SQ subtyping are shown in **Table 1**. In one embodiment, the gene expression levels of the classifier biomarkers for lung SQ subtyping are shown in **Table 1**. In one embodiment, all 80 genes can be used to classify the subtypes of SQ. In one embodiment, the first 20 genes are the selected gene signature biomarkers for basal, with gene numbers 1-10 up-regulated and gene numbers 11-20 down-regulated compared to a non-basal sample. In another embodiment, gene numbers 21-40 are the selected gene signature biomarkers specific for classical, with gene numbers 21-30 up-regulated and gene numbers 31-40 down-regulated compared to a non-classical sample. In yet another embodiment, gene numbers 41-60 are the selected gene signature biomarkers specific for primitive, with gene numbers 41-50 up-regulated and gene numbers 51-60 down-regulated compared to a non-primitive sample. In yet another embodiment, gene numbers 61-80 are the selected gene signature biomarkers specific for secretory, with gene numbers 61-70 up-regulated and gene numbers 71-80 down-regulated compared to a non-primitive sample.

**[0062] Table 1.** Gene Centroids of the 80 Classifier Biomarkers for the Lung Squamous Cell Carcinoma (SQ) Subtypes

| Gene No | Gene Symbol | Gene Name                                     | basal    | classical | primitive | secretory | GenBank Accession Number* | SEQ ID NO: |
|---------|-------------|-----------------------------------------------|----------|-----------|-----------|-----------|---------------------------|------------|
| 1       | SERP1NB4    | serpin family B member 4                      | 15.1924  | -1.28178  | -10.0199  | -7.32845  | NM_002974.3               | 1          |
| 2       | CXCL1       | C-X-C motif chemokine ligand 1                | 14.47981 | -8.31954  | -8.37503  | 0.217875  | NM_001511.3               | 2          |
| 3       | S100A9      | S100 calcium binding protein A9               | 14.35103 | -5.8793   | -9.10206  | -1.88807  | NM_002965.3               | 3          |
| 4       | S100A8      | S100 calcium binding protein A8               | 14.00816 | -4.229    | -9.53669  | -3.08348  | NM_001319196.1            | 4          |
| 5       | SERP1NB3    | serpin family B member 3                      | 13.97538 | 1.502713  | -10.9279  | -8.54433  | NM_006919.2               | 5          |
| 6       | EPHA2       | EPHA2                                         | 12.36835 | -4.75069  | -8.27087  | -1.67711  | NM_004431.4               | 6          |
| 7       | S100A2      | S100 calcium binding protein A2               | 12.02474 | 2.060853  | -9.93545  | -7.83677  | NM_005978.3               | 7          |
| 8       | MMP10       | matrix metalloproteinase 10                   | 11.70464 | -5.18263  | -3.79013  | -3.73457  | NM_002425.2               | 8          |
| 9       | IL4R        | interleukin 4 receptor                        | 11.67838 | -11.2637  | -9.61741  | 7.418712  | NM_000418.3               | 9          |
| 10      | PDZK1IP1    | PDZK1-interacting protein 1                   | 11.00384 | -9.67747  | -7.37829  | 4.707793  | NM_005764.3               | 10         |
| 11      | CDK5RAP2    | CDK5 regulatory subunit associated protein 2  | -13.3044 | 15.44094  | 0.582601  | -3.89079  | NM_018249.5               | 11         |
| 12      | FAM125B     | family with sequence similarity 125, member B | -12.2853 | 4.665284  | 4.308726  | 4.558947  | BC028675.1                | 12         |

|    |         |                                                    |          |          |          |          |                |    |
|----|---------|----------------------------------------------------|----------|----------|----------|----------|----------------|----|
| 13 | CABC1   | chaperone activity of bcl complex-like             | -10.3757 | 4.343061 | 7.391224 | 0.672574 | AB073905.1     | 13 |
| 14 | ODC1    | ornithine decarboxylase 1                          | -10.1908 | 15.84852 | 0.119301 | -7.30631 | NM_002539.2    | 14 |
| 15 | LPIN1   | lipin 1                                            | -10.134  | 3.748752 | 3.061368 | 4.230976 | NM_145693.2    | 15 |
| 16 | WASF1   | WAS protein family member 1                        | -9.89134 | 18.55734 | 1.814068 | -11.9252 | NM_003931.2    | 16 |
| 17 | USP13   | ubiquitin specific peptidase 13 (isopeptidase T-3) | -9.17202 | 7.072314 | 7.133335 | -3.50892 | NM_003940.2    | 17 |
| 18 | NUP210  | nucleoporin 210                                    | -8.91997 | 5.496247 | 2.508106 | 1.366756 | NM_024923.3    | 18 |
| 19 | GLI2    | GLI Family Zinc Finger 2                           | -8.58227 | 17.05556 | -5.643   | -6.1972  | NM_005270.4    | 19 |
| 20 | SPAG5   | sperm associated antigen 5                         | -8.26995 | 8.478108 | 6.146636 | -5.34162 | NM_006461.3    | 20 |
| 21 | ME1     | malic enzyme 1                                     | -11.1058 | 21.38387 | -2.66141 | -10.605  | NM_002395.5    | 21 |
| 22 | TALDO1  | transaldolase 1                                    | -11.3472 | 21.05835 | -2.95802 | -9.76549 | NM_006755.1    | 22 |
| 23 | AKR1C3  | aldo-keto reductase family 1, member C3            | -6.34178 | 19.62236 | -6.31166 | -10.9917 | NM_003739.5    | 23 |
| 24 | TXN     | thioredoxin                                        | -7.28934 | 19.56185 | -6.64144 | -9.68306 | NM_003329.3    | 24 |
| 25 | ALDH3A1 | aldehyde dehydrogenase 3 family member A1          | -4.42445 | 19.16675 | -7.69158 | -11.4995 | NM_001135168.1 | 25 |
| 26 | CHST7   | carbohydrate sulfotransferase 7                    | -6.70839 | 18.66004 | -5.80704 | -9.87835 | NM_019886.3    | 26 |
| 27 | ADAM23  | ADAM metallopeptidase domain 23                    | -7.14726 | 18.4093  | -5.05087 | -9.67848 | NM_003812.3    | 27 |

|    |         |                                                      |          |          |          |          |                |    |
|----|---------|------------------------------------------------------|----------|----------|----------|----------|----------------|----|
| 28 | TUFT1   | tuftelin 1                                           | -6.31534 | 18.07229 | -4.12497 | -10.8461 | NM_020127.2    | 28 |
| 29 | FOXE1   | forkhead box E1                                      | -2.047   | 17.53642 | -9.74136 | -10.6746 | NM_004473.3    | 29 |
| 30 | ALDH3A2 | aldehyde dehydrogenase 3 family member A2            | -7.7634  | 15.83759 | -4.12228 | -6.78263 | NM_001031806.1 | 30 |
| 31 | PHC2    | polyhomeotic homolog 2                               | 5.947711 | -19.3491 | 3.975339 | 12.79184 | NM_198040.2    | 31 |
| 32 | SLC43A3 | solute carrier family 43 member 3                    | 2.164732 | -15.4786 | 4.435501 | 12.06209 | NM_014096.3    | 32 |
| 33 | CAPZB   | capping actin protein of muscle Z-line beta subunit  | 9.697325 | -15.4337 | -0.08505 | 7.331941 | NM_004930.4    | 33 |
| 34 | FAM46A  | family with sequence similarity 46 member A          | 9.050488 | -14.8822 | 0.551123 | 6.928165 | NM_017633.2    | 34 |
| 35 | PTP4A2  | protein tyrosine phosphatase type IVA, member 2      | 5.400389 | -14.838  | 1.837093 | 9.801226 | NM_080391.3    | 35 |
| 36 | DPYD    | dihydropyrimidine dehydrogenase                      | 8.78203  | -14.5434 | -5.09695 | 10.92233 | NM_000110.3    | 36 |
| 37 | TRIM8   | tripartite motif containing 8                        | 3.847394 | -14.5393 | -1.94247 | 13.84298 | NM_030912.2    | 37 |
| 38 | CD47    | CD47 molecule                                        | 8.84354  | -14.3091 | -2.8533  | 8.964713 | NM_001777.3    | 38 |
| 39 | CRIP2   | cysteine rich protein 2                              | 4.809366 | -14.1729 | 1.781357 | 9.711258 | NM_001312.3    | 39 |
| 40 | ST3GAL5 | ST3 beta-galactosidase alpha-2,3-sialyltransferase 5 | 2.667885 | -13.865  | -1.29718 | 13.85595 | NM_003896.3    | 40 |

|    |             |                                                      |          |          |          |          |                    |    |
|----|-------------|------------------------------------------------------|----------|----------|----------|----------|--------------------|----|
| 41 | HSF2        | heat shock<br>transcripti<br>on factor 2             | -5.79001 | 1.050968 | 11.39169 | -3.33599 | NM_004506.<br>3    | 41 |
| 42 | MARCKSL1    | MARCKS<br>like 1                                     | 1.317716 | -10.696  | 9.825417 | 3.621776 | NM_023009.<br>6    | 42 |
| 43 | EFHD1       | EF-hand<br>domain<br>family<br>member<br>D1          | -2.47675 | -11.1247 | 9.620027 | 8.265181 | NM_025202.<br>3    | 43 |
| 44 | CHKA        | choline<br>kinase<br>alpha                           | -2.84869 | -7.08145 | 9.530024 | 4.135237 | NM_001277.<br>2    | 44 |
| 45 | PLEKH<br>B1 | pleckstrin<br>homology<br>domain<br>containing<br>B1 | -5.94374 | -6.54778 | 9.307835 | 6.960047 | NM_021200.<br>2    | 45 |
| 46 | FNBP1<br>L  | formin<br>binding<br>protein 1<br>like               | 2.207537 | -13.5657 | 9.226556 | 6.372445 | NM_0010249<br>48.2 | 46 |
| 47 | ZNF23<br>9  | zinc finger<br>protein<br>239                        | -2.61452 | -7.55963 | 8.698057 | 5.033708 | NM_005674.<br>2    | 47 |
| 48 | ABI2        | Abelson<br>interactor<br>2                           | -8.51982 | 0.375002 | 8.621929 | 2.322745 | NM_0012829<br>25.1 | 48 |
| 49 | MYL6<br>B   | Myosin<br>light chain<br>6B                          | -1.67839 | -4.74647 | 8.614632 | 0.913087 | NM_0011996<br>29.1 | 49 |
| 50 | TTLL4       | Tubulin<br>Tyrosine<br>Ligase<br>Like 4              | -4.42597 | -4.4529  | 8.316108 | 3.698664 | NM_014640.<br>4    | 50 |
| 51 | CLCA2       | Chloride<br>Channel<br>Accessory<br>2                | 11.3747  | 9.8531   | -13.5607 | -13.3641 | NM_006536.<br>5    | 51 |
| 52 | GJB3        | Gap<br>Junction<br>Protein<br>Beta 3                 | 9.738857 | 1.975392 | -12.8741 | -3.19459 | NM_024009.<br>2    | 52 |
| 53 | GPR87       | G Protein-<br>Coupled<br>Receptor<br>87              | 8.675319 | 3.714366 | -12.5406 | -4.28629 | NM_023915.<br>3    | 53 |
| 54 | SFN         | Stratifin                                            | 9.34036  | 7.030931 | -12.0548 | -9.10453 | NM_006142.<br>3    | 54 |
| 55 | CSTA        | Cystatin A                                           | 8.521125 | 6.642274 | -11.6462 | -8.09435 | NM_005213.<br>3    | 55 |
| 56 | DSG3        | Desmoglei<br>n 3                                     | 8.011909 | 9.629873 | -11.4831 | -11.0649 | NM_001944.<br>2    | 56 |

|    |            |                                                           |          |          |          |          |                |    |
|----|------------|-----------------------------------------------------------|----------|----------|----------|----------|----------------|----|
| 57 | ST6GALNAC2 | ST6 N-Acetylglucosaminidase Alpha-2,6-Sialyltransferase 2 | 3.15872  | 10.40711 | -11.4486 | -6.84553 | NM_006456.2    | 57 |
| 58 | GJB5       | Gap Junction Protein Beta 5                               | 9.68863  | 5.741838 | -11.4122 | -8.47546 | NM_005268.3    | 58 |
| 59 | TMPRSS4    | Transmembrane Protease, Serine 4                          | 7.421295 | 10.31518 | -10.907  | -11.6365 | NM_019894.3    | 59 |
| 60 | SDC1       | Syndecan 1                                                | 7.820035 | 8.717049 | -10.7889 | -10.3298 | NM_001006946.1 | 60 |
| 61 | FMNL1      | Formin Like 1                                             | -1.24826 | -12.3922 | -4.15625 | 18.39415 | NM_005892.3    | 61 |
| 62 | BIRC3      | Baculovirus IAP Repeat Containing 3                       | 0.52973  | -12.5421 | -4.71506 | 17.09129 | NM_001165.4    | 62 |
| 63 | ARHGDI1B   | Rho GDP Dissociation Inhibitor Beta                       | 1.579196 | -12.7865 | -4.70303 | 16.25141 | NM_001175.6    | 63 |
| 64 | SH2B3      | SH2B Adaptor Protein 3                                    | -3.48062 | -9.12196 | -3.04569 | 16.23607 | NM_005475.2    | 64 |
| 65 | HLA-DPA1   | Major Histocompatibility Complex, Class II, DP Alpha 1    | -2.12031 | -9.65989 | -3.99607 | 16.09867 | NM_033554.3    | 65 |
| 66 | NCF4       | Neutrophil Cytosolic Factor 4                             | 1.545361 | -11.6937 | -6.10253 | 16.0617  | NM_000631.4    | 66 |
| 67 | ACSL5      | Acyl-CoA Synthetase Long-Chain Family Member 5            | 1.654978 | -14.5012 | -1.66186 | 15.91216 | NM_016234.3    | 67 |
| 68 | CSF2RA     | Colony Stimulating Factor 2 Receptor Alpha                | -1.37456 | -10.508  | -2.90331 | 15.48108 | NM_006140.4    | 68 |

|    |              |                                                                                                                                |          |          |          |          |                    |    |
|----|--------------|--------------------------------------------------------------------------------------------------------------------------------|----------|----------|----------|----------|--------------------|----|
|    |              | Subunit                                                                                                                        |          |          |          |          |                    |    |
| 69 | LAPT<br>M5   | Lysosomal<br>Protein<br>Transmem-<br>brane 5                                                                                   | -1.16591 | -9.77656 | -4.28777 | 15.43442 | NM_006762.<br>2    | 69 |
| 70 | ARL6I<br>P5  | ADP-<br>ribosylatio<br>n-like<br>factor 6<br>interacting<br>protein 5                                                          | 3.195006 | -13.6479 | -4.55752 | 15.41665 | NM_006407.<br>3    | 70 |
| 71 | ADH7         | Alcohol<br>Dehydroge<br>nase 7<br>(Class IV),<br>Mu Or<br>Sigma                                                                | 0.182052 | 20.14673 | -9.26939 | -16.3334 | NM_0011665<br>04.1 | 71 |
| 72 | ABCC5        | ATP<br>Binding<br>Cassette<br>Subfamily<br>C Member<br>5                                                                       | -1.26645 | 17.73313 | -4.3337  | -15.6431 | NM_005688.<br>3    | 72 |
| 73 | SOX2         | SKY-Box<br>2                                                                                                                   | -2.70147 | 15.71135 | 0.455164 | -15.3051 | NM_003106.<br>3    | 73 |
| 74 | SLC9A<br>3R1 | Solute<br>Carrier<br>Family 9,<br>Subfamily<br>A (NHE3,<br>Cation<br>Proton<br>Antiporter<br>3),<br>Member 3<br>Regulator<br>1 | 1.902295 | 17.71886 | -9.60834 | -15.1497 | NM_004252.<br>4    | 74 |
| 75 | KLF5         | Kruppel-<br>Like<br>Factor 5<br>(Intestinal)                                                                                   | 4.456364 | 13.41893 | -8.16611 | -14.0138 | NM_001730.<br>4    | 75 |
| 76 | GPX2         | Glutathion<br>e<br>Peroxidase<br>2                                                                                             | -2.8397  | 17.49375 | -3.93026 | -14.0021 | NM_002083.<br>3    | 76 |
| 77 | PIR          | Pirin                                                                                                                          | -4.58676 | 16.97955 | -1.18296 | -13.5651 | NM_003662.<br>3    | 77 |
| 78 | TPD52<br>L1  | Tumor<br>Protein<br>DS2-Like                                                                                                   | 1.334706 | 10.49961 | 0.210322 | -13.4769 | NM_003287.<br>3    | 78 |

|    |        |                                    |          |          |          |          |             |    |
|----|--------|------------------------------------|----------|----------|----------|----------|-------------|----|
|    |        | 1                                  |          |          |          |          |             |    |
| 79 | SLC6A8 | Solute Carrier Family 6 Member 8   | 3.006892 | 11.83057 | -4.32575 | -13.4647 | NM_005629.3 | 79 |
| 80 | SIAH2  | SiAh E3 Ubiquitin Protein Ligase 2 | 1.897743 | 11.60785 | -2.92619 | -13.0552 | NM_005067.5 | 80 |

\*Each GenBank Accession Number is a representative or exemplary GenBank Accession Number for the listed gene and is herein incorporated by reference in its entirety for all purposes. Further, each listed representative or exemplary accession number should not be construed to limit the claims to the specific accession number.

**[0063] Table 2.** Classifier Biomarkers Selected for Basal, Classical, Primitive and Secretory SQ Subtypes

| <b>Basal</b> | <b>Classical</b> | <b>Primitive</b> | <b>Secretory</b> |
|--------------|------------------|------------------|------------------|
| SERPINB4     | ME1              | HSF2             | FMNL1            |
| CXCL1        | TALDO1           | MARCKSL1         | BIRC3            |
| S100A9       | AKR1C3           | EFHD1            | ARHGD1B          |
| S100A8       | TXN              | CHKA             | SH2B3            |
| SERPINB3     | ALDH3A1          | PLEKHB1          | HLA-DPA1         |
| EPHA2        | CHST7            | FBNP1L           | NCF4             |
| S100A2       | ADAM23           | ZNF239           | ACSL5            |
| MMP10        | TUFT1            | AB12             | CSF2RA           |
| IL4R         | FOXE1            | MYL6B            | LAPTM5           |
| PDZK1LP1     | ALDH3A2          | TTLL4            | ARL61P5          |
| CDK5RAP2     | PHC2             | CLCA2            | ADH7             |
| FAM125B      | SLC43A3          | GJB3             | ABCC5            |
| CABC1        | CAPZB            | GPR87            | SOX2             |
| CDC1         | FAM46A           | SFN              | SLC9A3R1         |
| LPIN1        | PTP4A2           | CSTA             | KLF5             |
| WASF1        | DPYD             | DSG3             | GPX2             |
| USP13        | TRIM8            | ST6GALNAC2       | PIR              |
| NUP210       | CD47             | GJB5             | TPD52L1          |
| GL12         | CRIP2            | TMPRSS4          | SLC6A8           |
| SPAG5        | ST3GAL5          | SDC1             | SIAH2            |

### **Diagnostic Uses**

[0064] In one embodiment, the methods and compositions provided herein allow for the differentiation of the four subtypes of squamous cell carcinoma: (1) basal; (2) classical; (3) primitive; and (4) secretory, with fewer genes needed than the molecular SQ subtyping methods known in the art.

[0065] In general, the methods provided herein are used to classify a lung cancer sample as a particular lung cancer subtype (e.g. subtype of squamous cell carcinoma). In one embodiment, the method comprises detecting or determining an expression level of at least one of the classifier biomarkers of any publically available Lung AD expression dataset. In one embodiment, the method comprises measuring, detecting or determining an expression level of at least one of the classifier biomarkers of **Table 1** in a lung cancer sample obtained from a patient or a subject.

[0066] The lung cancer sample for the detection or differentiation methods described herein can be a sample previously determined or diagnosed as a squamous cell carcinoma sample. The previous diagnosis can be based on a histological analysis. The histological analysis can be performed by one or more pathologists.

[0067] In one embodiment, the measuring or detecting step employed in the methods provided herein is at the nucleic acid level by performing RNA-seq, a reverse transcriptase polymerase chain reaction (RT-PCR) or a hybridization assay with oligonucleotides that are substantially complementary to portions of cDNA molecules of the at least one classifier biomarker (such as the classifier biomarkers of **Table 1**) under conditions suitable for RNA-seq, RT-PCR or hybridization and obtaining expression levels of the at least one classifier biomarkers based on the detecting step. The expression levels of the at least one of the classifier biomarkers are then compared to reference expression levels of the at least one of the classifier biomarker (such as the classifier biomarkers of **Table 1**) from at least one sample training set. The at least one sample training set can comprise, (i) expression levels of the at least one biomarker from a sample that overexpresses the at least one biomarker, (ii) expression levels from a reference basal, classical, primitive or secretory SQ subtype, or (iii) expression levels from an squamous cell carcinoma free lung sample, and classifying the lung tissue sample as a basal, classical, primitive or secretory subtype. The lung cancer sample can then be classified as a basal, classical, primitive or secretory subtype of squamous cell carcinoma based on the results of the comparing step. In one embodiment, the comparing step can comprise applying a statistical algorithm which comprises determining a correlation

between the expression data obtained from the lung tissue or cancer sample and the expression data from the at least one training set(s); and classifying the lung tissue or cancer sample as a basal, classical, primitive or secretory subtype based on the results of the statistical algorithm.

**[0068]** In one embodiment, the methods provided herein comprise probing the levels of at least one of the classifier biomarkers provided herein, such as the classifier biomarkers of **Table 1** at the nucleic acid level, in a lung cancer sample obtained from the patient. The lung cancer sample can be a sample previously determined or diagnosed as a squamous cell carcinoma sample. The previous diagnosis can be based on a histological analysis. The histological analysis can be performed by one or more pathologists. The probing step, in one embodiment, comprises mixing the sample with one or more oligonucleotides that are substantially complementary to portions of cDNA molecules of the at least one classifier biomarkers provided herein, such as the classifier biomarkers of **Table 1** under conditions suitable for hybridization of the one or more oligonucleotides to their complements or substantial complements; detecting whether hybridization occurs between the one or more oligonucleotides to their complements or substantial complements; and obtaining hybridization values of the at least one classifier biomarkers based on the detecting step. The hybridization values of the at least one classifier biomarkers are then compared to reference hybridization value(s) from at least one sample training set. For example, the at least one sample training set comprises hybridization values from a reference basal, classical, primitive or secretory sample. The lung cancer sample is classified, for example, as basal, classical, primitive or secretory based on the results of the comparing step.

**[0069]** The lung tissue sample can be any sample isolated from a human subject or patient. For example, in one embodiment, the analysis is performed on lung biopsies that are embedded in paraffin wax. In one embodiment, the sample can be a fresh frozen lung tissue sample. In another embodiment, the sample can be a bodily fluid obtained from the patient. The bodily fluid can be blood or fractions thereof (i.e., serum, plasma), urine, saliva, sputum or cerebrospinal fluid (CSF). The sample can contain cellular as well as extracellular sources of nucleic acid for use in the methods provided herein. The extracellular sources can be cell-free DNA and/or exosomes. In one embodiment, the sample can be a cell pellet or a wash. This aspect of the invention provides a means to improve current diagnostics by accurately identifying the major histological types, even from small biopsies. The methods of the

invention, including the RT-PCR methods, are sensitive, precise and have multi-analyte capability for use with paraffin embedded samples. See, for example, Cronin *et al.* (2004) *Am. J Pathol.* 164(1):35-42, herein incorporated by reference.

[0070] Formalin fixation and tissue embedding in paraffin wax is a universal approach for tissue processing prior to light microscopic evaluation. A major advantage afforded by formalin-fixed paraffin-embedded (FFPE) specimens is the preservation of cellular and architectural morphologic detail in tissue sections. (Fox et al. (1985) *J Histochem Cytochem* 33:845-853). The standard buffered formalin fixative in which biopsy specimens are processed is typically an aqueous solution containing 37% formaldehyde and 10-15% methyl alcohol. Formaldehyde is a highly reactive dipolar compound that results in the formation of protein-nucleic acid and protein-protein crosslinks in vitro (Clark et al. (1986) *J Histochem Cytochem* 34:1509-1512; McGhee and von Hippel (1975) *Biochemistry* 14:1281-1296, each incorporated by reference herein).

[0071] In one embodiment, the sample used herein is obtained from an individual, and comprises formalin-fixed paraffin-embedded (FFPE) tissue. However, other tissue and sample types are amenable for use herein. In one embodiment, the other tissue and sample types can be fresh frozen tissue, wash fluids, or cell pellets, or the like. In one embodiment, the sample can be a bodily fluid obtained from the individual. The bodily fluid can be blood or fractions thereof (e.g., serum, plasma), urine, sputum, saliva or cerebrospinal fluid (CSF). A biomarker nucleic acid as provided herein can be extracted from a cell or can be cell free or extracted from an extracellular vesicular entity such as an exosome.

[0072] Methods are known in the art for the isolation of RNA from FFPE tissue. In one embodiment, total RNA can be isolated from FFPE tissues as described by Bibikova et al. (2004) *American Journal of Pathology* 165:1799-1807, herein incorporated by reference. Likewise, the High Pure RNA Paraffin Kit (Roche) can be used. Paraffin is removed by xylene extraction followed by ethanol wash. RNA can be isolated from sectioned tissue blocks using the MasterPure Purification kit (Epicenter, Madison, Wis.); a DNase I treatment step is included. RNA can be extracted from frozen samples using Trizol reagent according to the supplier's instructions (Invitrogen Life Technologies, Carlsbad, Calif.). Samples with measurable residual genomic DNA can be resubjected to DNaseI treatment and assayed for DNA contamination. All purification, DNase treatment, and other steps can be performed

according to the manufacturer's protocol. After total RNA isolation, samples can be stored at -80 °C until use.

[0073] General methods for mRNA extraction are well known in the art and are disclosed in standard textbooks of molecular biology, including Ausubel et al., ed., Current Protocols in Molecular Biology, John Wiley & Sons, New York 1987-1999. Methods for RNA extraction from paraffin embedded tissues are disclosed, for example, in Rupp and Locker (Lab Invest. 56:A67, 1987) and De Andres et al. (Biotechniques 18:42-44, 1995). In particular, RNA isolation can be performed using a purification kit, a buffer set and protease from commercial manufacturers, such as Qiagen (Valencia, Calif.), according to the manufacturer's instructions. For example, total RNA from cells in culture can be isolated using Qiagen RNeasy mini-columns. Other commercially available RNA isolation kits include MasterPure™. Complete DNA and RNA Purification Kit (Epicentre, Madison, Wis.) and Paraffin Block RNA Isolation Kit (Ambion, Austin, Tex.). Total RNA from tissue samples can be isolated, for example, using RNA Stat-60 (Tel-Test, Friendswood, Tex.). RNA prepared from a tumor can be isolated, for example, by cesium chloride density gradient centrifugation. Additionally, large numbers of tissue samples can readily be processed using techniques well known to those of skill in the art, such as, for example, the single-step RNA isolation process of Chomczynski (U.S. Pat. No. 4,843,155, incorporated by reference in its entirety for all purposes).

[0074] In one embodiment, a sample comprises cells harvested from a lung tissue sample, for example, a squamous cell carcinoma sample. Cells can be harvested from a biological sample using standard techniques known in the art. For example, in one embodiment, cells are harvested by centrifuging a cell sample and resuspending the pelleted cells. The cells can be resuspended in a buffered solution such as phosphate-buffered saline (PBS). After centrifuging the cell suspension to obtain a cell pellet, the cells can be lysed to extract nucleic acid, *e.g.*, messenger RNA. All samples obtained from a subject, including those subjected to any sort of further processing, are considered to be obtained from the subject.

[0075] The sample, in one embodiment, is further processed before the detection of the biomarker levels of the combination of biomarkers set forth herein. For example, mRNA in a cell or tissue sample can be separated from other components of the sample. The sample can be concentrated and/or purified to isolate mRNA in its non-natural state, as the mRNA is not in its natural environment. For example, studies have indicated that the higher order structure

of mRNA *in vivo* differs from the *in vitro* structure of the same sequence (*see, e.g.,* Rouskin *et al.* (2014). *Nature* 505, pp. 701-705, incorporated herein in its entirety for all purposes).

[0076] mRNA from the sample in one embodiment, is hybridized to a synthetic DNA probe, which in some embodiments, includes a detection moiety (e.g., detectable label, capture sequence, barcode reporting sequence). Accordingly, in these embodiments, a non-natural mRNA-cDNA complex is ultimately made and used for detection of the biomarker. In another embodiment, mRNA from the sample is directly labeled with a detectable label, *e.g.,* a fluorophore. In a further embodiment, the non-natural labeled-mRNA molecule is hybridized to a cDNA probe and the complex is detected.

[0077] In one embodiment, once the mRNA is obtained from a sample, it is converted to complementary DNA (cDNA) in a hybridization reaction or is used in a hybridization reaction together with one or more cDNA probes. cDNA does not exist *in vivo* and therefore is a non-natural molecule. Furthermore, cDNA-mRNA hybrids are synthetic and do not exist *in vivo*. Besides cDNA not existing *in vivo*, cDNA is necessarily different than mRNA, as it includes deoxyribonucleic acid and not ribonucleic acid. The cDNA is then amplified, for example, by the polymerase chain reaction (PCR) or other amplification method known to those of ordinary skill in the art. For example, other amplification methods that may be employed include the ligase chain reaction (LCR) (Wu and Wallace, *Genomics*, 4:560 (1989), Landegren *et al.*, *Science*, 241:1077 (1988), incorporated by reference in its entirety for all purposes, transcription amplification (Kwoh *et al.*, *Proc. Natl. Acad. Sci. USA*, 86:1173 (1989), incorporated by reference in its entirety for all purposes), self-sustained sequence replication (Guatelli *et al.*, *Proc. Nat. Acad. Sci. USA*, 87:1874 (1990), incorporated by reference in its entirety for all purposes), incorporated by reference in its entirety for all purposes, and nucleic acid based sequence amplification (NASBA). Guidelines for selecting primers for PCR amplification are known to those of ordinary skill in the art. *See, e.g.,* McPherson *et al.*, *PCR Basics: From Background to Bench*, Springer-Verlag, 2000, incorporated by reference in its entirety for all purposes. The product of this amplification reaction, *i.e.*, amplified cDNA is also necessarily a non-natural product. First, as mentioned above, cDNA is a non-natural molecule. Second, in the case of PCR, the amplification process serves to create hundreds of millions of cDNA copies for every individual cDNA molecule of starting material. The numbers of copies generated are far removed from the number of copies of mRNA that are present *in vivo*.

[0078] In one embodiment, cDNA is amplified with primers that introduce an additional DNA sequence (e.g., adapter, reporter, capture sequence or moiety, barcode) onto the fragments (e.g., with the use of adapter-specific primers), or mRNA or cDNA biomarker sequences are hybridized directly to a cDNA probe comprising the additional sequence (e.g., adapter, reporter, capture sequence or moiety, barcode). Amplification and/or hybridization of mRNA to a cDNA probe therefore serves to create non-natural double stranded molecules from the non-natural single stranded cDNA, or the mRNA, by introducing additional sequences and forming non-natural hybrids. Further, as known to those of ordinary skill in the art, amplification procedures have error rates associated with them. Therefore, amplification introduces further modifications into the cDNA molecules. In one embodiment, during amplification with the adapter-specific primers, a detectable label, *e.g.*, a fluorophore, is added to single strand cDNA molecules. Amplification therefore also serves to create DNA complexes that do not occur in nature, at least because (i) cDNA does not exist *in vivo*, (i) adapter sequences are added to the ends of cDNA molecules to make DNA sequences that do not exist *in vivo*, (ii) the error rate associated with amplification further creates DNA sequences that do not exist *in vivo*, (iii) the disparate structure of the cDNA molecules as compared to what exists in nature, and (iv) the chemical addition of a detectable label to the cDNA molecules.

[0079] In some embodiments, the expression of a biomarker of interest is detected at the nucleic acid level via detection of non-natural cDNA molecules.

[0080] In some embodiments, the methods provided herein for lung cancer SQ subtyping includes detecting expression levels of a classifier biomarker set. In some embodiments, the detecting includes all of the classifier biomarkers of **Table 1** at the nucleic acid level or protein level. In another embodiment, a single or a subset or a plurality of the classifier biomarkers of **Table 1** are detected, for example, from about 10 to about 20. For example, in one embodiment, from about 5 to about 10, from about 10 to about 20, from about 20 to about 40, from about 40 to about 60, from about 60 to about 80 of the biomarkers in **Table 1** are detected in a method to determine the lung cancer SQ subtype. In another embodiment, each of the biomarkers from **Table 1** is detected in a method to determine the lung cancer subtype. In another embodiment, 20 of the biomarkers from **Table 1** are selected as the gene signatures for a specific lung cancer SQ subtype.

[0081] The detecting can be performed by any suitable technique including, but not limited to, RNA-seq, a reverse transcriptase polymerase chain reaction (RT-PCR), a microarray hybridization assay, or another hybridization assay, *e.g.*, a NanoString assay for example, with primers and/or probes specific to the classifier biomarkers, and/or the like. In some cases, the primers useful for the amplification methods (*e.g.*, RT-PCR or qRT-PCR) are any forward and reverse primers suitable for binding to a classifier gene provided herein, such as the classifier biomarkers listed in **Table 1**.

[0082] The biomarkers described herein include RNA comprising the entire or partial sequence of any of the nucleic acid sequences of interest, or their non-natural cDNA product, obtained synthetically *in vitro* in a reverse transcription reaction. The term “fragment” is intended to refer to a portion of the polynucleotide that generally comprise at least 10, 15, 20, 50, 75, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 800, 900, 1,000, 1,200, or 1,500 contiguous nucleotides, or up to the number of nucleotides present in a full-length biomarker polynucleotide disclosed herein. A fragment of a biomarker polynucleotide will generally encode at least 15, 25, 30, 50, 100, 150, 200, or 250 contiguous amino acids, or up to the total number of amino acids present in a full-length biomarker protein of the invention.

[0083] In some embodiments, overexpression, such as of an RNA transcript or its expression product, is determined by normalization to the level of reference RNA transcripts or their expression products, which can be all measured transcripts (or their products) in the sample or a particular reference set of RNA transcripts (or their non-natural cDNA products). Normalization is performed to correct for or normalize away both differences in the amount of RNA or cDNA assayed and variability in the quality of the RNA or cDNA used. Therefore, an assay typically measures and incorporates the expression of certain normalizing genes, including well known housekeeping genes, such as, for example, GAPDH and/or  $\beta$ -Actin. Alternatively, normalization can be based on the mean or median signal of all of the assayed biomarkers or a large subset thereof (global normalization approach).

[0084] Isolated mRNA can be used in hybridization or amplification assays that include, but are not limited to, Southern or Northern analyses, PCR analyses and probe arrays, NanoString Assays. One method for the detection of mRNA levels involves contacting the isolated mRNA or synthesized cDNA with a nucleic acid molecule (probe) that can hybridize to the mRNA encoded by the gene being detected. The nucleic acid probe can be, for example, a

cDNA, or a portion thereof, such as an oligonucleotide of at least 7, 15, 30, 50, 100, 250, or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to the non-natural cDNA or mRNA biomarker of the present invention.

[0085] As explained above, in one embodiment, once the mRNA is obtained from a sample, it is converted to complementary DNA (cDNA) in a hybridization reaction. Conversion of the mRNA to cDNA can be performed with oligonucleotides or primers comprising sequence that is complementary to a portion of a specific mRNA. Conversion of the mRNA to cDNA can be performed with oligonucleotides or primers comprising random sequence. Conversion of the mRNA to cDNA can be performed with oligonucleotides or primers comprising sequence that is complementary to the poly(A) tail of an mRNA. cDNA does not exist *in vivo* and therefore is a non-natural molecule. In a further embodiment, the cDNA is then amplified, for example, by the polymerase chain reaction (PCR) or other amplification method known to those of ordinary skill in the art. PCR can be performed with the forward and/or reverse primers comprising sequence complementary to at least a portion of a classifier gene provided herein, such as the classifier biomarkers in **Table 1**. The product of this amplification reaction, *i.e.*, amplified cDNA is necessarily a non-natural product. As mentioned above, cDNA is a non-natural molecule. Second, in the case of PCR, the amplification process serves to create hundreds of millions of cDNA copies for every individual cDNA molecule of starting material. The number of copies generated is far removed from the number of copies of mRNA that are present *in vivo*.

[0086] In one embodiment, cDNA is amplified with primers that introduce an additional DNA sequence (adapter sequence) onto the fragments (with the use of adapter-specific primers). The adaptor sequence can be a tail, wherein the tail sequence is not complementary to the cDNA. For example, the forward and/or reverse primers comprising sequence complementary to at least a portion of a classifier gene provided herein, such as the classifier biomarkers from **Table 1** can comprise tail sequence. Amplification therefore serves to create non-natural double stranded molecules from the non-natural single stranded cDNA, by introducing barcode, adapter and/or reporter sequences onto the already non-natural cDNA. In one embodiment, during amplification with the adapter-specific primers, a detectable label, *e.g.*, a fluorophore, is added to single strand cDNA molecules. Amplification therefore also serves to create DNA complexes that do not occur in nature, at least because (i) cDNA does not exist *in vivo*, (ii) adapter sequences are added to the ends of cDNA molecules to make

DNA sequences that do not exist *in vivo*, (iii) the error rate associated with amplification further creates DNA sequences that do not exist *in vivo*, (iv) the disparate structure of the cDNA molecules as compared to what exists in nature, and (v) the chemical addition of a detectable label to the cDNA molecules.

[0087] In one embodiment, the synthesized cDNA (for example, amplified cDNA) is immobilized on a solid surface via hybridization with a probe, e.g., via a microarray. In another embodiment, cDNA products are detected via real-time polymerase chain reaction (PCR) via the introduction of fluorescent probes that hybridize with the cDNA products. For example, in one embodiment, biomarker detection is assessed by quantitative fluorogenic RT-PCR (e.g., with TaqMan® probes). For PCR analysis, well known methods are available in the art for the determination of primer sequences for use in the analysis.

[0088] Biomarkers provided herein in one embodiment, are detected via a hybridization reaction that employs a capture probe and/or a reporter probe. For example, the hybridization probe is a probe derivatized to a solid surface such as a bead, glass or silicon substrate. In another embodiment, the capture probe is present in solution and mixed with the patient's sample, followed by attachment of the hybridization product to a surface, e.g., via a biotin-avidin interaction (e.g., where biotin is a part of the capture probe and avidin is on the surface). The hybridization assay, in one embodiment, employs both a capture probe and a reporter probe. The reporter probe can hybridize to either the capture probe or the biomarker nucleic acid. Reporter probes e.g., are then counted and detected to determine the level of biomarker(s) in the sample. The capture and/or reporter probe, in one embodiment contain a detectable label, and/or a group that allows functionalization to a surface.

[0089] For example, the nCounter gene analysis system (see, e.g., Geiss *et al.* (2008) Nat. Biotechnol. 26, pp. 317-325, incorporated by reference in its entirety for all purposes, is amenable for use with the methods provided herein.

[0090] Hybridization assays described in U.S. Patent Nos. 7,473,767 and 8,492,094, the disclosures of which are incorporated by reference in their entireties for all purposes, are amenable for use with the methods provided herein, i.e., to detect the biomarkers and biomarker combinations described herein.

[0091] Biomarker levels may be monitored using a membrane blot (such as used in hybridization analysis such as Northern, Southern, dot, and the like), or microwells, sample

tubes, gels, beads, or fibers (or any solid support comprising bound nucleic acids). See, for example, U.S. Pat. Nos. 5,770,722, 5,874,219, 5,744,305, 5,677,195 and 5,445,934, each incorporated by reference in their entireties.

**[0092]** In one embodiment, microarrays are used to detect biomarker levels. Microarrays are particularly well suited for this purpose because of the reproducibility between different experiments. DNA microarrays provide one method for the simultaneous measurement of the expression levels of large numbers of genes. Each array consists of a reproducible pattern of capture probes attached to a solid support. Labeled RNA or DNA is hybridized to complementary probes on the array and then detected by laser scanning hybridization intensities for each probe on the array are determined and converted to a quantitative value representing relative gene expression levels. See, for example, U.S. Pat. Nos. 6,040,138, 5,800,992 and 6,020,135, 6,033,860, and 6,344,316, each incorporated by reference in their entireties. High-density oligonucleotide arrays are particularly useful for determining the gene expression profile for a large number of RNAs in a sample.

**[0093]** Techniques for the synthesis of these arrays using mechanical synthesis methods are described in, for example, U.S. Pat. No. 5,384,261. Although a planar array surface is generally used, the array can be fabricated on a surface of virtually any shape or even a multiplicity of surfaces. Arrays can be nucleic acids (or peptides) on beads, gels, polymeric surfaces, fibers (such as fiber optics), glass, or any other appropriate substrate. See, for example, U.S. Pat. Nos. 5,770,358, 5,789,162, 5,708,153, 6,040,193 and 5,800,992, each incorporated by reference in their entireties. Arrays can be packaged in such a manner as to allow for diagnostics or other manipulation of an all-inclusive device. See, for example, U.S. Pat. Nos. 5,856,174 and 5,922,591, each incorporated by reference in their entireties.

**[0094]** Serial analysis of gene expression (SAGE) in one embodiment is employed in the methods described herein. SAGE is a method that allows the simultaneous and quantitative analysis of a large number of gene transcripts, without the need of providing an individual hybridization probe for each transcript. First, a short sequence tag (about 10-14 bp) is generated that contains sufficient information to uniquely identify a transcript, provided that the tag is obtained from a unique position within each transcript. Then, many transcripts are linked together to form long serial molecules, that can be sequenced, revealing the identity of the multiple tags simultaneously. The expression pattern of any population of transcripts can be quantitatively evaluated by determining the abundance of individual tags, and identifying

the gene corresponding to each tag. *See*, Velculescu et al. Science 270:484-87, 1995; Cell 88:243-51, 1997, incorporated by reference in its entirety.

**[0095]** An additional method of biomarker level analysis at the nucleic acid level is the use of a sequencing method, for example, RNAseq, next generation sequencing, and massively parallel signature sequencing (MPSS), as described by Brenner et al. (Nat. Biotech. 18:630-34, 2000, incorporated by reference in its entirety). This is a sequencing approach that combines non-gel-based signature sequencing with *in vitro* cloning of millions of templates on separate 5  $\mu\text{m}$  diameter microbeads. First, a microbead library of DNA templates is constructed by *in vitro* cloning. This is followed by the assembly of a planar array of the template-containing microbeads in a flow cell at a high density (typically greater than  $3.0 \times 10^6$  microbeads/cm<sup>2</sup>). The free ends of the cloned templates on each microbead are analyzed simultaneously, using a fluorescence-based signature sequencing method that does not require DNA fragment separation. This method has been shown to simultaneously and accurately provide, in a single operation, hundreds of thousands of gene signature sequences from a yeast cDNA library.

**[0096]** Another method of biomarker level expression analysis at the nucleic acid level is the use of an amplification method such as, for example, RT-PCR or quantitative RT-PCR (qRT-PCR). Methods for determining the level of biomarker mRNA in a sample may involve the process of nucleic acid amplification, e.g., by RT-PCR (the experimental embodiment set forth in Mullis, 1987, U.S. Pat. No. 4,683,202), ligase chain reaction (Barany (1991) Proc. Natl. Acad. Sci. USA 88:189-193), self-sustained sequence replication (Guatelli et al. (1990) Proc. Natl. Acad. Sci. USA 87:1874-1878), transcriptional amplification system (Kwoh et al. (1989) Proc. Natl. Acad. Sci. USA 86:1173-1177), Q-Beta Replicase (Lizardi et al. (1988) Bio/Technology 6:1197), rolling circle replication (Lizardi et al., U.S. Pat. No. 5,854,033) or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques well known to those of skill in the art. Numerous different PCR or qRT-PCR protocols are known in the art and can be directly applied or adapted for use using the presently described compositions for the detection and/or quantification of expression of discriminative genes in a sample. *See*, for example, Fan et al. (2004) Genome Res. 14:878-885, herein incorporated by reference. Generally, in PCR, a target polynucleotide sequence is amplified by reaction with at least one oligonucleotide primer or pair of oligonucleotide primers. The primer(s) hybridize to a complementary region of the

target nucleic acid and a DNA polymerase extends the primer(s) to amplify the target sequence. Under conditions sufficient to provide polymerase-based nucleic acid amplification products, a nucleic acid fragment of one size dominates the reaction products (the target polynucleotide sequence which is the amplification product). The amplification cycle is repeated to increase the concentration of the single target polynucleotide sequence. The reaction can be performed in any thermocycler commonly used for PCR.

[0097] Quantitative RT-PCR (qRT-PCR) (also referred as real-time RT-PCR) is preferred under some circumstances because it provides not only a quantitative measurement, but also reduced time and contamination. As used herein, "quantitative PCR" (or "real time qRT-PCR") refers to the direct monitoring of the progress of a PCR amplification as it is occurring without the need for repeated sampling of the reaction products. In quantitative PCR, the reaction products may be monitored via a signaling mechanism (e.g., fluorescence) as they are generated and are tracked after the signal rises above a background level but before the reaction reaches a plateau. The number of cycles required to achieve a detectable or "threshold" level of fluorescence varies directly with the concentration of amplifiable targets at the beginning of the PCR process, enabling a measure of signal intensity to provide a measure of the amount of target nucleic acid in a sample in real time. A DNA binding dye (e.g., SYBR green) or a labeled probe can be used to detect the extension product generated by PCR amplification. Any probe format utilizing a labeled probe comprising the sequences of the invention may be used.

[0098] Immunohistochemistry methods are also suitable for detecting the levels of the biomarkers of the present invention. Samples can be frozen for later preparation or immediately placed in a fixative solution. Tissue samples can be fixed by treatment with a reagent, such as formalin, gluteraldehyde, methanol, or the like and embedded in paraffin. Methods for preparing slides for immunohistochemical analysis from formalin-fixed, paraffin-embedded tissue samples are well known in the art.

[0099] In one embodiment, the levels of the biomarkers provided herein, such as the classifier biomarkers of **Table 1** (or subsets thereof, for example 10 to 20, 20 to 30, 30 to 40, 40 to 50, 50 to 60, 60 to 70, or 70 to 80 biomarkers), are normalized against the expression levels of all RNA transcripts or their non-natural cDNA expression products, or protein products in the sample, or of a reference set of RNA transcripts or a reference set of their

non-natural cDNA expression products, or a reference set of their protein products in the sample.

**[00100]** In one embodiment, lung squamous cell carcinoma subtypes can be evaluated using levels of protein expression of one or more of the classifier genes provided herein, such as the classifier biomarkers listed in **Table 1**. The level of protein expression can be measured using an immunological detection method. Immunological detection methods which can be used herein include, but are not limited to, competitive and non-competitive assay systems using techniques such as Western blots, radioimmunoassays, ELISA (enzyme linked immunosorbent assay), "sandwich" immunoassays, immunoprecipitation assays, precipitin reactions, gel diffusion precipitin reactions, immunodiffusion assays, agglutination assays, complement-fixation assays, immunoradiometric assays, fluorescent immunoassays, protein A immunoassays, and the like. Such assays are routine and well known in the art (see, e.g., Ausubel *et al*, eds, 1994, *Current Protocols in Molecular Biology*, Vol. I, John Wiley & Sons, Inc., New York, which is incorporated by reference herein in its entirety).

**[00101]** In one embodiment, antibodies specific for biomarker proteins are utilized to detect the expression of a biomarker protein in a body sample. The method comprises obtaining a body sample from a patient or a subject, contacting the body sample with at least one antibody directed to a biomarker that is selectively expressed in lung cancer cells, and detecting antibody binding to determine if the biomarker is expressed in the patient sample. A preferred aspect of the present invention provides an immunocytochemistry technique for diagnosing lung cancer subtypes. One of skill in the art will recognize that the immunocytochemistry method described herein below may be performed manually or in an automated fashion.

**[00102]** As provided throughout, the methods set forth herein provide a method for determining the lung cancer SQ subtype of a patient. Once the biomarker levels are determined, for example by measuring non-natural cDNA biomarker levels or non-natural mRNA-cDNA biomarker complexes, the biomarker levels are compared to reference values or a reference sample, for example with the use of statistical methods or direct comparison of detected levels, to make a determination of the lung cancer molecular SQ subtype. Based on the comparison, the patient's lung cancer sample is SQ classified, e.g., as basal, classical, primitive or secretory.

**[00103]** In one embodiment, expression level values of the at least one classifier biomarkers provided herein, such as the classifier biomarkers of **Table 1** are compared to reference expression level value(s) from at least one sample training set, wherein the at least one sample training set comprises expression level values from a reference sample(s). In a further embodiment, the at least one sample training set comprises expression level values of the at least one classifier biomarkers provided herein, such as the classifier biomarkers of **Table 1** from a basal, classical, primitive, secretory sample, or a combination thereof.

**[00104]** In a separate embodiment, hybridization values of the at least one classifier biomarkers provided herein, such as the classifier biomarkers of **Table 1** are compared to reference hybridization value(s) from at least one sample training set, wherein the at least one sample training set comprises hybridization values from a reference sample(s). In a further embodiment, the at least one sample training set comprises hybridization values of the at least one classifier biomarkers provided herein, such as the classifier biomarkers of **Table 1** from a proximal basal, classical, primitive, secretory sample, or a combination thereof. Methods for comparing detected levels of biomarkers to reference values and/or reference samples are provided herein. Based on this comparison, in one embodiment a correlation between the biomarker levels obtained from the subject's sample and the reference values is obtained. An assessment of the lung cancer SQ subtype is then made.

**[00105]** Various statistical methods can be used to aid in the comparison of the biomarker levels obtained from the patient and reference biomarker levels, for example, from at least one sample training set.

**[00106]** In one embodiment, a supervised pattern recognition method is employed. Examples of supervised pattern recognition methods can include, but are not limited to, the nearest centroid methods (Dabney (2005) *Bioinformatics* 21(22):4148-4154 and Tibshirani et al. (2002) *Proc. Natl. Acad. Sci. USA* 99(10):6576-6572); soft independent modeling of class analysis (SIMCA) (see, for example, Wold, 1976); partial least squares analysis (PLS) (see, for example, Wold, 1966; Joreskog, 1982; Frank, 1984; Bro, R., 1997); linear discriminant analysis (LDA) (see, for example, Nilsson, 1965); K-nearest neighbour analysis (KNN) (see, for example, Brown et al., 1996); artificial neural networks (ANN) (see, for example, Wasserman, 1989; Anker et al., 1992; Hare, 1994); probabilistic neural networks (PNNs) (see, for example, Parzen, 1962; Bishop, 1995; Speck, 1990; Broomhead et al., 1988; Patterson, 1996); rule induction (RI) (see, for example, Quinlan, 1986); and, Bayesian

methods (see, for example, Bretthorst, 1990a, 1990b, 1988). In one embodiment, the classifier for identifying tumor subtypes based on gene expression data is the centroid based method described in Mullins et al. (2007) Clin Chem. 53(7):1273-9, each of which is herein incorporated by reference in its entirety.

**[00107]** In other embodiments, an unsupervised training approach is employed, and therefore, no training set is used.

**[00108]** Referring to sample training sets for supervised learning approaches again, in some embodiments, a sample training set(s) can include expression data of a plurality or all of the classifier biomarkers (e.g., all the classifier biomarkers of **Table 1**) from an squamous cell carcinoma sample. The plurality of classifier biomarkers can comprise at least two classifier biomarkers, at least 10 classifier biomarkers, at least 20 classifier biomarkers, at least 30 classifier biomarkers, at least 40 classifier biomarkers, at least 50 classifier biomarkers, at least 60 classifier biomarkers, or at least 70 classifier biomarkers of **Table 1**. In some embodiments, the sample training set(s) are normalized to remove sample-to-sample variation.

**[00109]** In some embodiments, comparing can include applying a statistical algorithm, such as, for example, any suitable multivariate statistical analysis model, which can be parametric or non-parametric. In some embodiments, applying the statistical algorithm can include determining a correlation between the expression data obtained from the human lung tissue sample and the expression data from the squamous cell carcinoma training set(s). In some embodiments, cross-validation is performed, such as (for example), leave-one-out cross-validation (LOOCV). In some embodiments, integrative correlation is performed. In some embodiments, a Spearman correlation is performed. In some embodiments, a centroid based method is employed for the statistical algorithm as described in Mullins *et al.* (2007) Clin Chem. 53(7):1273-9, and based on gene expression data, which is herein incorporated by reference in its entirety.

**[00110]** Results of the gene expression performed on a sample from a subject (test sample) may be compared to a biological sample(s) or data derived from a biological sample(s) that is known or suspected to be normal (“reference sample” or “normal sample”, e.g., non- squamous cell carcinoma sample). In some embodiments, a reference sample or reference gene expression data is obtained or derived from an individual known to have a

particular molecular subtype of squamous cell carcinoma, *i.e.*, basal, classical, secretory or primitive.

[00111] The reference sample may be assayed at the same time, or at a different time from the test sample. Alternatively, the biomarker level information from a reference sample may be stored in a database or other means for access at a later date.

[00112] The biomarker level results of an assay on the test sample may be compared to the results of the same assay on a reference sample. In some cases, the results of the assay on the reference sample are from a database, or a reference value(s). In some cases, the results of the assay on the reference sample are a known or generally accepted value or range of values by those skilled in the art. In some cases the comparison is qualitative. In other cases the comparison is quantitative. In some cases, qualitative or quantitative comparisons may involve but are not limited to one or more of the following: comparing fluorescence values, spot intensities, absorbance values, chemiluminescent signals, histograms, critical threshold values, statistical significance values, expression levels of the genes described herein, mRNA copy numbers.

[00113] In one embodiment, an odds ratio (OR) is calculated for each biomarker level panel measurement. Here, the OR is a measure of association between the measured biomarker values for the patient and an outcome, *e.g.*, lung squamous cell carcinoma subtype. For example, see, *J. Can. Acad. Child Adolesc. Psychiatry* 2010; 19(3): 227-229, which is incorporated by reference in its entirety for all purposes.

[00114] In one embodiment, a specified statistical confidence level may be determined in order to provide a confidence level regarding the lung cancer subtype. For example, it may be determined that a confidence level of greater than 90% may be a useful predictor of the lung cancer subtype. In other embodiments, more or less stringent confidence levels may be chosen. For example, a confidence level of about or at least about 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 97.5%, 99%, 99.5%, or 99.9% may be chosen. The confidence level provided may in some cases be related to the quality of the sample, the quality of the data, the quality of the analysis, the specific methods used, and/or the number of gene expression values (*i.e.*, the number of genes) analyzed. The specified confidence level for providing the likelihood of response may be chosen on the basis of the expected number of false positives or false negatives. Methods for choosing parameters for achieving a specified confidence

level or for identifying markers with diagnostic power include but are not limited to Receiver Operating Characteristic (ROC) curve analysis, binormal ROC, principal component analysis, odds ratio analysis, partial least squares analysis, singular value decomposition, least absolute shrinkage and selection operator analysis, least angle regression, and the threshold gradient directed regularization method.

**[00115]** Determining the lung squamous cell carcinoma subtype in some cases can be improved through the application of algorithms designed to normalize and or improve the reliability of the gene expression data. In some embodiments of the present invention, the data analysis utilizes a computer or other device, machine or apparatus for application of the various algorithms described herein due to the large number of individual data points that are processed. A “machine learning algorithm” refers to a computational-based prediction methodology, also known to persons skilled in the art as a “classifier,” employed for characterizing a gene expression profile or profiles, *e.g.*, to determine the lung squamous cell carcinoma subtype. The biomarker levels, determined by, *e.g.*, microarray-based hybridization assays, sequencing assays (*e.g.*, RNAseq), NanoString assays, quantitative amplification assays, etc., are in one embodiment subjected to the algorithm in order to classify the profile. Supervised learning generally involves “training” a classifier to recognize the distinctions among subtypes such as basal positive, classical positive, secretory positive or primitive positive, and then “testing” the accuracy of the classifier on an independent test set. Therefore, for new, unknown samples the classifier can be used to predict, for example, the class (*e.g.*, basal vs. classical vs. secretory vs. magnoid) in which the samples belong.

**[00116]** In some embodiments, a robust multi-array average (RMA) method may be used to normalize raw data. The RMA method begins by computing background-corrected intensities for each matched cell on a number of microarrays. In one embodiment, the background corrected values are restricted to positive values as described by Irizarry *et al.* (2003). *Biostatistics* April 4 (2): 249-64, incorporated by reference in its entirety for all purposes. After background correction, the base-2 logarithm of each background corrected matched-cell intensity is then obtained. The background corrected, log-transformed, matched intensity on each microarray is then normalized using the quantile normalization method in which for each input array and each probe value, the array percentile probe value is replaced with the average of all array percentile points, this method is more completely described by Bolstad *et al.* *Bioinformatics* 2003, incorporated by reference in its entirety. Following

quantile normalization, the normalized data may then be fit to a linear model to obtain an intensity measure for each probe on each microarray. Tukey's median polish algorithm (Tukey, J. W., *Exploratory Data Analysis*. 1977, incorporated by reference in its entirety for all purposes) may then be used to determine the log-scale intensity level for the normalized probe set data.

[00117] Various other software programs may be implemented. In certain methods, feature selection and model estimation may be performed by logistic regression with *lasso* penalty using *glmnet* (Friedman *et al.* (2010). *Journal of statistical software* 33(1): 1-22, incorporated by reference in its entirety). Raw reads may be aligned using TopHat (Trapnell *et al.* (2009). *Bioinformatics* 25(9): 1105-11, incorporated by reference in its entirety). In methods, top features (N ranging from 10 to 200) are used to train a linear support vector machine (SVM) (Suykens JAK, Vandewalle J. Least Squares Support Vector Machine Classifiers. *Neural Processing Letters* 1999; 9(3): 293-300, incorporated by reference in its entirety) using the *e1071* library (Meyer D. Support vector machines: the interface to libsvm in package e1071. 2014, incorporated by reference in its entirety). Confidence intervals, in one embodiment, are computed using the pROC package (Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC bioinformatics* 2011; 12: 77, incorporated by reference in its entirety).

[00118] In addition, data may be filtered to remove data that may be considered suspect. In one embodiment, data derived from microarray probes that have fewer than about 4, 5, 6, 7 or 8 guanosine + cytosine nucleotides may be considered to be unreliable due to their aberrant hybridization propensity or secondary structure issues. Similarly, data deriving from microarray probes that have more than about 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, or 22 guanosine + cytosine nucleotides may in one embodiment be considered unreliable due to their aberrant hybridization propensity or secondary structure issues.

[00119] In some embodiments of the present invention, data from probe-sets may be excluded from analysis if they are not identified at a detectable level (above background).

[00120] In some embodiments of the present disclosure, probe-sets that exhibit no, or low variance may be excluded from further analysis. Low-variance probe-sets are excluded from the analysis via a Chi-Square test. In one embodiment, a probe-set is considered to be low-variance if its transformed variance is to the left of the 99 percent confidence interval of the

Chi-Squared distribution with  $(N-1)$  degrees of freedom.  $(N-1) \times \text{Probe-set Variance} / (\text{Gene Probe-set Variance})$ .  $\text{Chi-Sq}(N-1)$  where  $N$  is the number of input CEL files,  $(N-1)$  is the degrees of freedom for the Chi-Squared distribution, and the “probe-set variance for the gene” is the average of probe-set variances across the gene. In some embodiments of the present invention, probe-sets for a given mRNA or group of mRNAs may be excluded from further analysis if they contain less than a minimum number of probes that pass through the previously described filter steps for GC content, reliability, variance and the like. For example in some embodiments, probe-sets for a given gene or transcript cluster may be excluded from further analysis if they contain less than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, or less than about 20 probes.

**[00121]** Methods of biomarker level data analysis in one embodiment further include the use of a feature selection algorithm as provided herein. In some embodiments of the present invention, feature selection is provided by use of the LIMMA software package (Smyth, G. K. (2005). Limma: linear models for microarray data. In: Bioinformatics and Computational Biology Solutions using R and Bioconductor, R. Gentleman, V. Carey, S. Dudoit, R. Irizarry, W. Huber (eds.), Springer, New York, pages 397-420, incorporated by reference in its entirety for all purposes).

**[00122]** Methods of biomarker level data analysis, in one embodiment, include the use of a pre-classifier algorithm. For example, an algorithm may use a specific molecular fingerprint to pre-classify the samples according to their composition and then apply a correction/normalization factor. This data/information may then be fed in to a final classification algorithm which would incorporate that information to aid in the final diagnosis.

**[00123]** Methods of biomarker level data analysis, in one embodiment, further include the use of a classifier algorithm as provided herein. In one embodiment of the present invention, a diagonal linear discriminant analysis, k-nearest neighbor algorithm, support vector machine (SVM) algorithm, linear support vector machine, random forest algorithm, or a probabilistic model-based method or a combination thereof is provided for classification of microarray data. In some embodiments, identified markers that distinguish samples (*e.g.*, of varying biomarker level profiles, and/or varying molecular subtypes of squamous cell carcinoma (*e.g.*, basal, classical, secretory, primitive)) are selected based on statistical significance of the difference in biomarker levels between classes of interest. In some cases, the statistical

significance is adjusted by applying a Benjamin Hochberg or another correction for false discovery rate (FDR).

**[00124]** In some cases, the classifier algorithm may be supplemented with a meta-analysis approach such as that described by Fishel and Kaufman et al. 2007 *Bioinformatics* 23(13): 1599-606, incorporated by reference in its entirety for all purposes. In some cases, the classifier algorithm may be supplemented with a meta-analysis approach such as a repeatability analysis.

**[00125]** Methods for deriving and applying posterior probabilities to the analysis of biomarker level data are known in the art and have been described for example in Smyth, G. K. 2004 *Stat. Appi. Genet. Mol. Biol.* 3: Article 3, incorporated by reference in its entirety for all purposes. In some cases, the posterior probabilities may be used in the methods of the present invention to rank the markers provided by the classifier algorithm.

**[00126]** A statistical evaluation of the results of the biomarker level profiling may provide a quantitative value or values indicative of one or more of the following: molecular subtype of squamous cell carcinoma (basal, classical, secretory, primitive); the likelihood of the success of a particular therapeutic intervention, *e.g.*, angiogenesis inhibitor therapy, chemotherapy, or immunotherapy. In one embodiment, the data is presented directly to the physician in its most useful form to guide patient care, or is used to define patient populations in clinical trials or a patient population for a given medication. The results of the molecular profiling can be statistically evaluated using a number of methods known to the art including, but not limited to: the students T test, the two sided T test, Pearson rank sum analysis, hidden Markov model analysis, analysis of q-q plots, principal component analysis, one way ANOVA, two way ANOVA, LIMMA and the like.

**[00127]** In some cases, accuracy may be determined by tracking the subject over time to determine the accuracy of the original diagnosis. In other cases, accuracy may be established in a deterministic manner or using statistical methods. For example, receiver operator characteristic (ROC) analysis may be used to determine the optimal assay parameters to achieve a specific level of accuracy, specificity, positive predictive value, negative predictive value, and/or false discovery rate.

**[00128]** In some cases, the results of the biomarker level profiling assays, are entered into a database for access by representatives or agents of a molecular profiling business, the

individual, a medical provider, or insurance provider. In some cases, assay results include sample classification, identification, or diagnosis by a representative, agent or consultant of the business, such as a medical professional. In other cases, a computer or algorithmic analysis of the data is provided automatically. In some cases the molecular profiling business may bill the individual, insurance provider, medical provider, researcher, or government entity for one or more of the following: molecular profiling assays performed, consulting services, data analysis, reporting of results, or database access.

**[00129]** In some embodiments of the present invention, the results of the biomarker level profiling assays are presented as a report on a computer screen or as a paper record. In some embodiments, the report may include, but is not limited to, such information as one or more of the following: the levels of biomarkers (*e.g.*, as reported by copy number or fluorescence intensity, etc.) as compared to the reference sample or reference value(s); the likelihood the subject will respond to a particular therapy, based on the biomarker level values and the lung squamous cell carcinoma subtype and proposed therapies.

**[00130]** In one embodiment, the results of the gene expression profiling may be classified into one or more of the following: basal positive, classical positive, secretory positive, primitive positive, basal negative, classical negative, secretory negative, primitive negative; likely to respond to angiogenesis inhibitor, immunotherapy or chemotherapy; unlikely to respond to angiogenesis inhibitor, immunotherapy or chemotherapy; or a combination thereof.

**[00131]** In some embodiments of the present invention, results are classified using a trained algorithm. Trained algorithms of the present invention include algorithms that have been developed using a reference set of known gene expression values and/or normal samples, for example, samples from individuals diagnosed with a particular molecular subtype of squamous cell carcinoma. In some cases, a reference set of known gene expression values are obtained from individuals who have been diagnosed with a particular molecular subtype of squamous cell carcinoma, and are also known to respond (or not respond) to angiogenesis inhibitor therapy. In some cases, a reference set of known gene expression values are obtained from individuals who have been diagnosed with a particular molecular subtype of squamous cell carcinoma, and are also known to respond (or not respond) to immunotherapy. In some cases, a reference set of known gene expression values are obtained from individuals who have been diagnosed with a particular molecular subtype

of squamous cell carcinoma, and are also known to respond (or not respond) to chemotherapy.

**[00132]** Algorithms suitable for categorization of samples include but are not limited to k-nearest neighbor algorithms, support vector machines, linear discriminant analysis, diagonal linear discriminant analysis, updown, naive Bayesian algorithms, neural network algorithms, hidden Markov model algorithms, genetic algorithms, or any combination thereof.

**[00133]** When a binary classifier is compared with actual true values (*e.g.*, values from a biological sample), there are typically four possible outcomes. If the outcome from a prediction is p (where “p” is a positive classifier output, such as the presence of a deletion or duplication syndrome) and the actual value is also p, then it is called a true positive (TP); however if the actual value is n then it is said to be a false positive (FP). Conversely, a true negative has occurred when both the prediction outcome and the actual value are n (where “n” is a negative classifier output, such as no deletion or duplication syndrome), and false negative is when the prediction outcome is n while the actual value is p. In one embodiment, consider a test that seeks to determine whether a person is likely or unlikely to respond to angiogenesis inhibitor therapy. A false positive in this case occurs when the person tests positive, but actually does respond. A false negative, on the other hand, occurs when the person tests negative, suggesting they are unlikely to respond, when they actually are likely to respond. The same holds true for classifying a lung cancer subtype.

**[00134]** The positive predictive value (PPV), or precision rate, or post-test probability of disease, is the proportion of subjects with positive test results who are correctly diagnosed as likely or unlikely to respond, or diagnosed with the correct lung cancer subtype, or a combination thereof. It reflects the probability that a positive test reflects the underlying condition being tested for. Its value does however depend on the prevalence of the disease, which may vary. In one example the following characteristics are provided: FP (false positive); TN (true negative); TP (true positive); FN (false negative). False positive rate ( $\square$ )=FP/(FP+TN)-specificity; False negative rate ( $\square$ )=FN/(TP+FN)-sensitivity; Power=sensitivity = 1- $\square$  $\square$ ; Likelihood-ratio positive=sensitivity/(1-specificity); Likelihood-ratio negative=(1-sensitivity)/specificity. The negative predictive value (NPV) is the proportion of subjects with negative test results who are correctly diagnosed.

[00135] In some embodiments, the results of the biomarker level analysis of the subject methods provide a statistical confidence level that a given diagnosis is correct. In some embodiments, such statistical confidence level is at least about, or more than about 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 99.5%, or more.

[00136] In some embodiments, the method further includes classifying the lung tissue sample as a particular lung cancer subtype based on the comparison of biomarker levels in the sample and reference biomarker levels, for example present in at least one training set. In some embodiments, the lung tissue sample is classified as a particular subtype if the results of the comparison meet one or more criterion such as, for example, a minimum percent agreement, a value of a statistic calculated based on the percentage agreement such as (for example) a kappa statistic, a minimum correlation (*e.g.*, Pearson's correlation) and/or the like.

[00137] It is intended that the methods described herein can be performed by software (stored in memory and/or executed on hardware), hardware, or a combination thereof. Hardware modules may include, for example, a general-purpose processor, a field programmable gate array (FPGA), and/or an application specific integrated circuit (ASIC). Software modules (executed on hardware) can be expressed in a variety of software languages (*e.g.*, computer code), including Unix utilities, C, C++, Java™, Ruby, SQL, SAS®, the R programming language/software environment, Visual Basic™, and other object-oriented, procedural, or other programming language and development tools. Examples of computer code include, but are not limited to, micro-code or micro-instructions, machine instructions, such as produced by a compiler, code used to produce a web service, and files containing higher-level instructions that are executed by a computer using an interpreter. Additional examples of computer code include, but are not limited to, control signals, encrypted code, and compressed code.

[00138] Some embodiments described herein relate to devices with a non-transitory computer-readable medium (also can be referred to as a non-transitory processor-readable medium or memory) having instructions or computer code thereon for performing various computer-implemented operations and/or methods disclosed herein. The computer-readable medium (or processor-readable medium) is non-transitory in the sense that it does not include transitory propagating signals per se (*e.g.*, a propagating electromagnetic wave carrying information on a transmission medium such as space or a cable). The media and computer

code (also can be referred to as code) may be those designed and constructed for the specific purpose or purposes. Examples of non-transitory computer-readable media include, but are not limited to: magnetic storage media such as hard disks, floppy disks, and magnetic tape; optical storage media such as Compact Disc/Digital Video Discs (CD/DVDs), Compact Disc-Read Only Memories (CD-ROMs), and holographic devices; magneto-optical storage media such as optical disks; carrier wave signal processing modules; and hardware devices that are specially configured to store and execute program code, such as Application-Specific Integrated Circuits (ASICs), Programmable Logic Devices (PLDs), Read-Only Memory (ROM) and Random-Access Memory (RAM) devices. Other embodiments described herein relate to a computer program product, which can include, for example, the instructions and/or computer code discussed herein.

**[00139]** In some embodiments, a single biomarker, or from about 10 to about 20, from about 20 to about 30, from about 20 to about 40, from about 40 to about 50, from about 40 to about 60, from about 60 to about 70, from about 60 to about 80, from about 20 to about 60, from about 20 to about 80, from about 40 to about 80 biomarkers (e.g., as disclosed in **Table 1**) is capable of classifying subtypes of lung squamous cell carcinoma with a predictive success of at least about 70%, at least about 71%, at least about 72%, about 73%, about 74%, about 75%, about 76%, about 77%, about 78%, about 79%, about 80%, about 81%, about 82%, about 83%, about 84%, about 85%, about 86%, about 87%, about 88%, about 89%, about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, about 99%, up to 100%, and all values in between. In some embodiments, any combination of biomarkers disclosed herein (e.g., in **Table 1**) can be used to obtain a predictive success of at least about 70%, at least about 71%, at least about 72%, about 73%, about 74%, about 75%, about 76%, about 77%, about 78%, about 79%, about 80%, about 81%, about 82%, about 83%, about 84%, about 85%, about 86%, about 87%, about 88%, about 89%, about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, about 99%, up to 100%, and all values in between.

**[00140]** In some embodiments, a single biomarker, or from about 10 to about 20, from about 20 to about 30, from about 20 to about 40, from about 40 to about 50, from about 40 to about 60, from about 60 to about 70, from about 60 to about 80, from about 20 to about 60, from about 20 to about 80, from about 40 to about 80 biomarkers (e.g., as disclosed in **Table 1**) is capable of classifying lung squamous cell carcinoma subtypes with a sensitivity or

specificity of at least about 70%, at least about 71%, at least about 72%, about 73%, about 74%, about 75%, about 76%, about 77%, about 78%, about 79%, about 80%, about 81%, about 82%, about 83%, about 84%, about 85%, about 86%, about 87%, about 88%, about 89%, about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, about 99%, up to 100%, and all values in between. In some embodiments, any combination of biomarkers disclosed herein can be used to obtain a sensitivity or specificity of at least about 70%, at least about 71%, at least about 72%, about 73%, about 74%, about 75%, about 76%, about 77%, about 78%, about 79%, about 80%, about 81%, about 82%, about 83%, about 84%, about 85%, about 86%, about 87%, about 88%, about 89%, about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, about 99%, up to 100%, and all values in between.

### **Classifier Gene Selection**

**[00141]** In one embodiment, the methods and compositions provided herein are useful for analyzing the expression of a set of biomarkers in a sample (e.g., lung tissue sample or a lung SQ sample) from a patient, whereby the set of biomarkers comprise a fewer number of biomarkers than methods known in the art for molecularly classifying lung SQ subtype. In some cases, the set of biomarkers is less than 250, 240, 230, 220, 210, 200, 150, 100, 95 or 90 biomarkers. In some cases, the set of biomarkers is the set of 80 biomarkers listed in **Table 1**. In some cases, the set of biomarkers is a sub-set of biomarkers listed **Table 1**. The biomarkers or classifier genes useful in the methods and compositions provided herein can be selected from one or more lung squamous cell carcinoma datasets from one or more databases. The databases can be public databases. In one embodiment, classifier genes (e.g., one or more genes listed in **Table 1** and **Table 2**) useful in the methods and compositions provided herein for detecting or diagnosing lung squamous cell carcinoma subtypes were selected from a lung squamous cell carcinoma RNAseq dataset from The Cancer Genome Atlas (TCGA). In one embodiment, classifier genes useful for the methods and compositions provided herein such as those in **Table 1** are selected by subjecting a large set of classifier genes to an *in silico* based process in order to determine the minimum number of genes whose expression profile can be used to determine an SQ subtype of sample obtained from a subject. In some cases, the large set of classifier genes can be a lung SQ RNAseq dataset such as, for example, from TCGA. In some cases, the large set of classifier genes can be the 208-gene classifier disclosed in Wilkerson et al. (Clin Cancer Res 2010; 16(19):4864-4875),

whereby the 208-gene classifier can serve to define gold standard subtype. The *in silico* process for selecting a gene cassette as provided herein for determining lung SQ subtype of a sample from a patient can comprise, applying or using a Classifying arrays to Nearest Centroid (CLaNC) algorithm with modification on the standard 208 classifier genes to choose an equal number of negatively and positively correlated genes for each subtype. For determination of the optimal number of genes (e.g, 20 per subtype as shown in **Table 1**) to include in the signature, the process can further comprise performing a 5-fold cross validation using TCGA lung squamous cell carcinoma dataset as provided herein to produce cross-validation curves as shown in **FIG. 9**. To get the final list of gene classifiers, the method can further comprise applying the Classifying arrays to Nearest Centroid (CLaNC) to the entire TCGA data set minus 20% of samples with the lowest gold standard subtype prediction strength, and removing an equal number from each subtype.

**[00142]** In one embodiment, the method further comprises validating the gene classifiers. Validation can comprise testing the expression of the classifiers in several fresh frozen publicly available array and RNAseq datasets and calling the subtype based on said expression levels and subsequently comparing the expression with the gold standard subtype calls as defined by the previously published 208-gene signature disclosed in Wilkerson et al. (Clin Cancer Res 2010; 16(19):4864-4875). Final validation of the gene signature (e.g., **Table 1**) can then be performed in a newly collected RNAseq dataset of archived formalin-fixed paraffin-embedded (FFPE) squamous cell carcinoma samples to assure comparable performance in the FFPE samples. In one embodiment, the classifier biomarkers of **Table 1** were selected based on the *in silico* CLaNC process described herein. The gene symbols and official gene names are listed in column 2 and column 3, respectively.

**[00143]** In one embodiment, the methods of the invention require the detection of at least 1, 2, 3, 4, 5, 6, 7 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79 or 80 classifier biomarkers in a lung cancer cell sample (e.g., lung SQ cancer sample) obtained from a patient in order to identify a basal, classical, secretory or primitive lung squamous cell carcinoma subtype. The same applies for other classifier gene expression datasets as provided herein.

[00144] In another embodiment, the methods of the invention require the detection of a total of at least 1, at least 2, at least 5, at least 10, at least 20, at least 40, at least 60 or up to 80 classifier biomarkers out of the 80 gene biomarkers of **Table 1** in a lung cancer cell sample (e.g., lung SQ cancer sample) obtained from a patient in order to identify a basal, classical, secretory or primitive lung squamous cell carcinoma subtype. The same applies for other classifier gene expression datasets as provided herein.

[00145] In one embodiment, at least 1, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9 or up to 10 biomarkers of **Table 1** are “up-regulated” in a specific subtype of lung squamous cell carcinoma. In another embodiment, at least 1, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9 or up to 10 biomarkers of **Table 1** are “down-regulated” in a specific subtype of lung squamous cell carcinoma. The same applies for other classifier gene expression datasets as provided herein.

[00146] In one embodiment, the expression level of an “up-regulated” biomarker as provided herein is increased by about 0.5-fold, about 1-fold, about 1.5-fold, about 2-fold, about 2.5-fold, about 3-fold, about 3.5-fold, about 4-fold, about 4.5-fold, about 5-fold, and any values in between. In another embodiment, the expression level of a “down-regulated” biomarker as provided herein is decreased by about 0.8-fold, about 1.4-fold, about 2-fold, about 2.6-fold, about 3.2-fold, about 3.6-fold, about 4-fold, and any values in between.

[00147] It is recognized that additional genes or proteins can be used in the practice of the invention. In general, genes useful in classifying the subtypes of lung squamous cell carcinoma, include those that are independently capable of distinguishing between normal versus tumor, or between different classes or grades of lung cancer. A gene is considered to be capable of reliably distinguishing between subtypes if the area under the receiver operator characteristic (ROC) curve is approximately 1.

### **Clinical / Therapeutic Uses**

[00148] In one embodiment, a method is provided herein for determining a disease outcome or prognosis for a patient suffering from cancer. In some cases, the cancer is lung cancer. The disease outcome or prognosis can be measured by examining the overall survival for a period of time or intervals (e.g., 0 to 36 months or 0 to 60 months). In one embodiment, survival is analyzed as a function of subtype (e.g., for lung cancer, SQ (basal, classical,

secretory or primitive)). Relapse-free and overall survival can be assessed using standard Kaplan-Meier plots as well as Cox proportional hazards modeling.

[00149] In one embodiment, upon determining a patient's lung cancer subtype, the patient is selected for suitable therapy, for example chemotherapy or drug therapy with an angiogenesis inhibitor or immunotherapy. In one embodiment, upon determining a patient's lung cancer subtype, the patient is administered a suitable therapeutic agent, for example chemotherapeutic agent(s) or an angiogenesis inhibitor or immunotherapeutic agent(s). In one embodiment, the therapy is immunotherapy, and the immunotherapeutic agent is a checkpoint inhibitor, monoclonal antibody, biological response modifier, therapeutic vaccine or cellular immunotherapy.

[00150] The methods of present invention are also useful for evaluating clinical response to therapy, as well as for endpoints in clinical trials for efficacy of new therapies. The extent to which sequential diagnostic expression profiles move towards normal can be used as one measure of the efficacy of the candidate therapy.

[00151] In one embodiment, the methods of the invention also find use in predicting response to different lines of therapies based on the subtype of lung squamous cell carcinoma (SQ). For example, chemotherapeutic response can be improved by more accurately assigning tumor subtypes. Likewise, treatment regimens can be formulated based on the tumor subtype. For example, clinical trials have shown convincing evidence that the VEGF inhibitor, bevacizumab, can be effective in the treatment of NSCLC. In one embodiment, the primitive SQ subtype can have enhanced response to immunotherapy. In another embodiment, all subtypes can have enhanced response to chemotherapies, angiogenesis inhibitor treatments, and immunotherapies.

[00152] In one embodiment, upon determining a patient's lung squamous cell carcinoma subtype, the patient is selected for suitable therapy, for example chemotherapy, immunotherapy or drug therapy with an angiogenesis inhibitor. In one embodiment, upon determining a patient's lung squamous cell carcinoma subtype using the methods provided herein, a suitable therapeutic agent, for example a chemotherapeutic agent(s), an immunotherapeutic agent or an angiogenesis inhibitor is administered to the patient.

### **Angiogenesis Inhibitors**

**[00153]** In one embodiment, upon determining a patient's lung SQ subtype, the patient is selected for drug therapy with an angiogenesis inhibitor. Upon making a determination of whether a patient is likely to respond to angiogenesis inhibitor therapy, or selecting a patient for angiogenesis inhibitor therapy, in one embodiment, the patient is administered the angiogenesis inhibitor. The angiogenesis inhibitor can be any of the angiogenesis inhibitors described herein

**[00154]** In one embodiment, the therapy is angiogenesis inhibitor therapy, and the angiogenesis inhibitor is a vascular endothelial growth factor (VEGF) inhibitor, a VEGF receptor inhibitor, a platelet derived growth factor (PDGF) inhibitor or a PDGF receptor inhibitor.

**[00155]** Each biomarker panel can include one, two, three, four, five, six, seven, eight, nine, ten, 20, 40, 60, 80 or more biomarkers usable by a classifier (also referred to as a "classifier biomarker") to assess whether an squamous cell carcinoma patient is likely to respond to angiogenesis inhibitor therapy; to select an squamous cell carcinoma patient for angiogenesis inhibitor therapy; to determine a "hypoxia score" and/or to subtype an squamous cell carcinoma sample as basal, classical, secretory or primitive molecular subtype. As used herein, the term "classifier" can refer to any algorithm for statistical classification, and can be implemented in hardware, in software, or a combination thereof. The classifier can be capable of 2-level, 3-level, 4-level, or higher, classification, and can depend on the nature of the entity being classified. One or more classifiers can be employed to achieve the aspects disclosed herein.

**[00156]** In general, methods of determining whether a squamous cell carcinoma patient is likely to respond to angiogenesis inhibitor therapy, or methods of selecting an squamous cell carcinoma patient for angiogenesis inhibitor therapy are provided herein. In one embodiment, the method comprises assessing whether the patient's squamous cell carcinoma subtype is basal, classical, secretory or primitive using the methods described herein (e.g., assessing the expression of one or more classifier biomarkers of **Table 1**) and probing an squamous cell carcinoma sample from the patient for the levels of at least five biomarkers selected from the group consisting of RRAGD, FABP5, UCHL1, GAL, PLOD, DDIT4, VEGF, ADM, ANGPTL4, NDRG1, NP, SLC16A3, and C14ORF58 (see **Table 3**) at the nucleic acid level. In a further embodiment, the probing step comprises mixing the sample with five or more oligonucleotides that are substantially complementary to portions of nucleic

acid molecules of the at least five biomarkers under conditions suitable for hybridization of the five or more oligonucleotides to their complements or substantial complements, detecting whether hybridization occurs between the five or more oligonucleotides to their complements or substantial complements; and obtaining hybridization values of the sample based on the detecting steps. The hybridization values of the sample are then compared to reference hybridization value(s) from at least one sample training set, wherein the at least one sample training set comprises (i) hybridization value(s) of the at least five biomarkers from a sample that overexpresses the at least five biomarkers, or overexpresses a subset of the at least five biomarkers, (ii) hybridization values of the at least five biomarkers from a reference basal, classical, secretory or primitive sample, or (iii) hybridization values of the at least five biomarkers from an squamous cell carcinoma free lung sample. A determination of whether the patient is likely to respond to angiogenesis inhibitor therapy, or a selection of the patient for angiogenesis inhibitor is then made based upon (i) the patient's squamous cell carcinoma subtype and (ii) the results of comparison.

| <b>Table 3. Biomarkers for hypoxia profile</b> |                                                                     |                               |
|------------------------------------------------|---------------------------------------------------------------------|-------------------------------|
| <b>Name</b>                                    | <b>Abbreviation</b>                                                 | <b>GenBank Accession No.*</b> |
| RRAGD                                          | Ras-related GTP binding D                                           | BC003088                      |
| FABP5                                          | fatty acid binding protein 5                                        | M94856                        |
| UCHL1                                          | ubiquitin carboxyl-terminal esterase L1                             | NM_004181                     |
| GAL                                            | Galanin                                                             | BC030241                      |
| PLOD                                           | procollagen-lysine, 2-oxoglutarate 5-dioxygenase lysine hydroxylase | M98252                        |
| DDIT4                                          | DNA-damage-inducible transcript 4                                   | NM_019058                     |
| VEGF                                           | vascular endothelial growth factor                                  | M32977                        |
| ADM                                            | Adrenomedullin                                                      | NM_001124                     |
| ANGPTL4                                        | angiopoietin-like 4                                                 | AF202636                      |
| NDRG1                                          | N-myc downstream regulated gene 1                                   | NM_006096                     |
| NP                                             | nucleoside phosphorylase                                            | NM_000270                     |
| SLC16A3                                        | solute carrier family 16 monocarboxylic acid transporters, member 3 | NM_004207                     |
| C14ORF58                                       | chromosome 14 open reading frame 58                                 | AK000378                      |

\*Each GenBank Accession Number is a representative or exemplary GenBank Accession Number for the listed gene and is herein incorporated by reference in its entirety for all

purposes. Further, each listed representative or exemplary accession number should not be construed to limit the claims to the specific accession number.

**[00157]** The aforementioned set of thirteen biomarkers, or a subset thereof, is also referred to herein as a “hypoxia profile”.

**[00158]** In one embodiment, the method provided herein includes determining the levels of at least five biomarkers, at least six biomarkers, at least seven biomarkers, at least eight biomarkers, at least nine biomarkers, or at least ten biomarkers, or five to thirteen, six to thirteen, seven to thirteen, eight to thirteen, nine to thirteen or ten to thirteen biomarkers selected from RRAGD, FABP5, UCHL1, GAL, PLOD, DDIT4, VEGF, ADM, ANGPTL4, NDRG1, NP, SLC16A3, and C14ORF58 in an squamous cell carcinoma sample obtained from a subject. Biomarker expression in some instances may be normalized against the expression levels of all RNA transcripts or their expression products in the sample, or against a reference set of RNA transcripts or their expression products. The reference set as explained throughout, may be an actual sample that is tested in parallel with the squamous cell carcinoma sample, or may be a reference set of values from a database or stored dataset. Levels of expression, in one embodiment, are reported in number of copies, relative fluorescence value or detected fluorescence value. The level of expression of the biomarkers of the hypoxia profile together with squamous cell carcinoma subtype as determined using the methods provided herein can be used in the methods described herein to determine whether a patient is likely to respond to angiogenesis inhibitor therapy.

**[00159]** In one embodiment, the levels of expression of the thirteen biomarkers (or subsets thereof, as described above, *e.g.*, five or more, from about five to about 13), are normalized against the expression levels of all RNA transcripts or their non-natural cDNA expression products, or protein products in the sample, or of a reference set of RNA transcripts or a reference set of their non-natural cDNA expression products, or a reference set of their protein products in the sample.

**[00160]** In one embodiment, angiogenesis inhibitor treatments include, but are not limited to an integrin antagonist, a selectin antagonist, an adhesion molecule antagonist, an antagonist of intercellular adhesion molecule (ICAM)-1, ICAM-2, ICAM-3, platelet endothelial adhesion molecule (PCAM), vascular cell adhesion molecule (VCAM), lymphocyte function-associated antigen 1 (LFA-1), a basic fibroblast growth factor

antagonist, a vascular endothelial growth factor (VEGF) modulator, a platelet derived growth factor (PDGF) modulator (*e.g.*, a PDGF antagonist).

**[00161]** In one embodiment of determining whether a subject is likely to respond to an integrin antagonist, the integrin antagonist is a small molecule integrin antagonist, for example, an antagonist described by Paolillo *et al.* (Mini Rev Med Chem, 2009, volume 12, pp. 1439-1446, incorporated by reference in its entirety), or a leukocyte adhesion-inducing cytokine or growth factor antagonist (*e.g.*, tumor necrosis factor- $\alpha$  (TNF- $\alpha$ ), interleukin-1 $\beta$  (IL-1 $\beta$ ), monocyte chemotactic protein-1 (MCP-1) and a vascular endothelial growth factor (VEGF)), as described in U.S. Patent No. 6,524,581, incorporated by reference in its entirety herein.

**[00162]** The methods provided herein are also useful for determining whether a subject is likely to respond to one or more of the following angiogenesis inhibitors: interferon gamma 1 $\beta$ , interferon gamma 1 $\beta$  (Actimmune®) with pirfenidone, ACUHTR028,  $\alpha$ V $\beta$ 5, aminobenzoate potassium, amyloid P, ANG1122, ANG1170, ANG3062, ANG3281, ANG3298, ANG4011, anti-CTGF RNAi, Aplidin, astragalus membranaceus extract with salvia and schisandra chinensis, atherosclerotic plaque blocker, Azol, AZX100, BB3, connective tissue growth factor antibody, CT140, danazol, Esbriet, EXC001, EXC002, EXC003, EXC004, EXC005, F647, FG3019, Fibrocorin, Follistatin, FT011, a galectin-3 inhibitor, GKT137831, GMCT01, GMCT02, GRMD01, GRMD02, GRN510, Heberon Alfa R, interferon  $\alpha$ -2 $\beta$ , ITMN520, JKB119, JKB121, JKB122, KRX168, LPA1 receptor antagonist, MGN4220, MIA2, microRNA 29a oligonucleotide, MMI0100, noscapine, PBI4050, PBI4419, PDGFR inhibitor, PF-06473871, PGN0052, Pirespa, Pirfenex, pirfenidone, plitidepsin, PRM151, Px102, PYN17, PYN22 with PYN17, Relivergen, rhPTX2 fusion protein, RXI109, secretin, STX100, TGF- $\beta$  Inhibitor, transforming growth factor,  $\beta$ -receptor 2 oligonucleotide, VA999260, XV615 or a combination thereof.

**[00163]** In another embodiment, a method is provided for determining whether a subject is likely to respond to one or more endogenous angiogenesis inhibitors. In a further embodiment, the endogenous angiogenesis inhibitor is endostatin, a 20 kDa C-terminal fragment derived from type XVIII collagen, angiostatin (a 38 kDa fragment of plasmin), a member of the thrombospondin (TSP) family of proteins. In a further embodiment, the angiogenesis inhibitor is a TSP-1, TSP-2, TSP-3, TSP-4 and TSP-5. Methods for determining the likelihood of response to one or more of the following angiogenesis

inhibitors are also provided a soluble VEGF receptor, *e.g.*, soluble VEGFR-1 and neuropilin 1 (NPR1), angiopoietin-1, angiopoietin-2, vasostatin, calreticulin, platelet factor-4, a tissue inhibitor of metalloproteinase (TIMP) (*e.g.*, TIMP1, TIMP2, TIMP3, TIMP4), cartilage-derived angiogenesis inhibitor (*e.g.*, peptide troponin I and chondromodulin I), a disintegrin and metalloproteinase with thrombospondin motif 1, an interferon (IFN), (*e.g.*, IFN- $\alpha$ , IFN- $\beta$ , IFN- $\gamma$ ), a chemokine, *e.g.*, a chemokine having the C-X-C motif (*e.g.*, CXCL10, also known as interferon gamma-induced protein 10 or small inducible cytokine B10), an interleukin cytokine (*e.g.*, IL-4, IL-12, IL-18), prothrombin, antithrombin III fragment, prolactin, the protein encoded by the *TNFSF15* gene, osteopontin, maspin, canstatin, proliferin-related protein.

**[00164]** In one embodiment, a method for determining the likelihood of response to one or more of the following angiogenesis inhibitors is provided is angiopoietin-1, angiopoietin-2, angiostatin, endostatin, vasostatin, thrombospondin, calreticulin, platelet factor-4, TIMP, CDAI, interferon  $\alpha$ , interferon  $\beta$ , vascular endothelial growth factor inhibitor (VEGI) meth-1, meth-2, prolactin, VEGI, SPARC, osteopontin, maspin, canstatin, proliferin-related protein (PRP), restin, TSP-1, TSP-2, interferon gamma 1 $\beta$ , ACUHTR028,  $\alpha$ V $\beta$ 5, aminobenzoate potassium, amyloid P, ANG1122, ANG1170, ANG3062, ANG3281, ANG3298, ANG4011, anti-CTGF RNAi, Aplidin, astragalus membranaceus extract with salvia and schisandra chinensis, atherosclerotic plaque blocker, Azol, AZX100, BB3, connective tissue growth factor antibody, CT140, danazol, Esbriet, EXC001, EXC002, EXC003, EXC004, EXC005, F647, FG3019, Fibrocorin, Follistatin, FT011, a galectin-3 inhibitor, GKT137831, GMCT01, GMCT02, GRMD01, GRMD02, GRN510, Heberon Alfa R, interferon  $\alpha$ -2 $\beta$ , ITMN520, JKB119, JKB121, JKB122, KRX168, LPA1 receptor antagonist, MGN4220, MIA2, microRNA 29a oligonucleotide, MMI0100, noscapine, PBI4050, PBI4419, PDGFR inhibitor, PF-06473871, PGN0052, Pirespa, Pirfenex, pirfenidone, plitidepsin, PRM151, Px102, PYN17, PYN22 with PYN17, Relivergen, rhPTX2 fusion protein, RXI109, secretin, STX100, TGF- $\beta$  Inhibitor, transforming growth factor,  $\beta$ -receptor 2 oligonucleotide, VA999260, XV615 or a combination thereof.

**[00165]** In yet another embodiment, the angiogenesis inhibitor can include pazopanib (Votrient), sunitinib (Sutent), sorafenib (Nexavar), axitinib (Inlyta), ponatinib (Iclusig), vandetanib (Caprelsa), cabozantinib (Cometrig), ramucirumab (Cyramza), regorafenib (Stivarga), ziv-aflibercept (Zaltrap), motesanib, or a combination thereof. In another

embodiment, the angiogenesis inhibitor is a VEGF inhibitor. In a further embodiment, the VEGF inhibitor is axitinib, cabozantinib, aflibercept, brivanib, tivozanib, ramucirumab or motesanib. In yet a further embodiment, the angiogenesis inhibitor is motesanib.

**[00166]** In one embodiment, the methods provided herein relate to determining a subject's likelihood of response to an antagonist of a member of the platelet derived growth factor (PDGF) family, for example, a drug that inhibits, reduces or modulates the signaling and/or activity of PDGF-receptors (PDGFR). For example, the PDGF antagonist, in one embodiment, is an anti-PDGF aptamer, an anti-PDGF antibody or fragment thereof, an anti-PDGFR antibody or fragment thereof, or a small molecule antagonist. In one embodiment, the PDGF antagonist is an antagonist of the PDGFR- $\alpha$  or PDGFR- $\beta$ . In one embodiment, the PDGF antagonist is the anti-PDGF- $\beta$  aptamer E10030, sunitinib, axitinib, sorefenib, imatinib, imatinib mesylate, nintedanib, pazopanib HCl, ponatinib, MK-2461, dovitinib, pazopanib, crenolanib, PP-121, telatinib, imatinib, KRN 633, CP 673451, TSU-68, Ki8751, amuvatinib, tivozanib, masitinib, motesanib diphosphate, dovitinib dilactic acid, linifanib (ABT-869).

### **Immunotherapy**

**[00167]** In one embodiment, provided herein is a method for determining whether a squamous cell carcinoma (SQ) lung cancer patient is likely to respond to immunotherapy by determining the subtype of SQ of a sample obtained from the patient and, based on the SQ lung cancer subtype, assessing whether the patient is likely to respond to immunotherapy. In another embodiment, provided herein is a method of selecting a patient suffering from SQ for immunotherapy by determining an SQ subtype of a sample from the patient and, based on the SQ subtype, selecting the patient for immunotherapy. The determination of the SQ subtype of the sample obtained from the patient can be performed using any method for subtyping SQ known in the art. In one embodiment, the sample obtained from the patient has been previously diagnosed as being SQ, and the methods provided herein are used to determine the SQ subtype of the sample. The previous diagnosis can be based on a histological analysis. The histological analysis can be performed by one or more pathologists. In one embodiment, the SQ subtyping is performed via gene expression analysis of a set or panel of biomarkers or subsets thereof in order to generate an expression profile. The gene expression analysis can be performed on a lung cancer sample (e.g., lung cancer SQ sample) obtained from a patient in order to determine the presence, absence or level of expression of one or more biomarkers selected from a publically available lung cancer database described herein and/or **Table 1**

provided herein. The SQ subtype can be selected from the group consisting of primitive, classical, secretory and basal. The immunotherapy can be any immunotherapy provided herein. In one embodiment, the immunotherapy comprises administering one or more checkpoint inhibitors. The checkpoint inhibitors can be any checkpoint inhibitor provided herein such as, for example, a checkpoint inhibitor that targets PD-1, PD-L1 or CTLA4.

**[00168]** As disclosed herein, the biomarkers panels, or subsets thereof, can be those disclosed in any publically available SQ gene expression dataset or datasets. In one embodiment, the lung cancer is SQ and the biomarker panel or subset thereof is, for example, TCGA lung SQ RNAseq gene expression dataset (n=501). In one embodiment, the lung cancer is SQ and the biomarker panel or subset thereof is, for example, the SQ gene expression dataset (n=75) disclosed in Lee et al. (Cancer Res 2008; 14(22): 7397-7404), the contents of which are herein incorporated by reference in its entirety. In one embodiment, the lung cancer is SQ and the biomarker panel or subset thereof is, for example, the SQ gene expression dataset (n=130) disclosed in Raponi et al. (Cancer Res 2006; 66(7): 466-472), the contents of which are herein incorporated by reference in its entirety. In one embodiment, the lung cancer is SQ and the biomarker panel or subset thereof is, for example, the SQ gene expression dataset (n=56) disclosed in Wilkerson et al. (Clin Cancer Res 2010; 16(19):4864-4875), the contents of which are herein incorporated by reference in its entirety. In one embodiment, the lung cancer is SQ and the biomarker panel or subset thereof is, for example, the SQ gene expression dataset disclosed in **Table 1**. In **Table 2**, the first column of the table represents the biomarker list for distinguishing basal. The second column of the table represents the biomarker list for classical. The third column of the table represents the biomarker list for distinguishing primitive. The last column of the table represents the biomarker list for distinguishing secretory. In some cases, as shown in **Table 2**, a total of 80 biomarkers can be used for SQ subtype determination. For each SQ subtype in **Table 2**, 10 of the 20 biomarkers can be negatively correlated genes, while 10 can be positively correlated genes which can be selected as the gene signature of a specific SQ subtype.

**[00169]** In some embodiments, the method for lung cancer subtyping (e.g., SQ subtyping) includes detecting expression levels of a classifier biomarker set. The classifier biomarker set can be a set of biomarkers from a publically available database such as, for example, TCGA lung SQ RNASeq gene expression dataset(s) or any other dataset provided

herein. In some embodiments, the detecting includes all of the classifier biomarkers of **Table 1** or any other dataset provided herein at the nucleic acid level or protein level. In another embodiment, a single classifier biomarker of **Table 1** or a subset of the classifier biomarkers of **Table 1** or any other dataset provided herein are detected, for example, from about five to about twenty. In another embodiment, a single classifier biomarker of **Table 1** or a subset of the classifier biomarkers of **Table 1** and/or any other dataset provided herein are detected, for example, from about 20 to about 80. In another embodiment, all of the classifier biomarkers of **Table 1** or any other dataset provided herein are detected. In another embodiment, at least one or all of the classifier biomarkers of **Table 1** in combination with one or more classifier biomarkers of any other SQ dataset provided herein are detected. The detecting can be performed by any suitable technique including, but not limited to, RNA-seq, a reverse transcriptase polymerase chain reaction (RT-PCR), a microarray hybridization assay, or another hybridization assay, *e.g.*, a NanoString assay for example, with primers and/or probes specific to the classifier biomarkers, and/or the like. In some cases, the primers useful for the amplification methods (*e.g.*, RT-PCR or qRT-PCR) are any forward and reverse primers suitable for binding to a classifier gene from a dataset provided herein alone or in combination.

**[00170]** In one embodiment, from about 1 to about 5, about 5 to about 10, from about 5 to about 15, from about 5 to about 20, from about 5 to about 25, from about 5 to about 30, from about 5 to about 35, from about 5 to about 40, from about 5 to about 45, from about 5 to about 50, from about 5 to about 55, from about 5 to about 60, from about 5 to about 65, from about 5 to about 70, from about 5 to about 75, or from about 5 to about 80 of the biomarkers in any of the SQ gene expression datasets provided herein, including, for example, **Table 1** for an SQ lung sample are detected in a method to determine the lung cancer subtype as provided herein. In another embodiment, each of the biomarkers from any one of the SQ gene expression datasets provided herein, including, for example, **Table 1** for an SQ lung sample are detected in a method to determine the lung cancer subtype as provided herein.

**[00171]** In one embodiment, the methods provided herein further comprise determining the presence, absence or level of immune activation in a SQ subtype. The presence or level of immune cell activation can be determined by creating an expression profile or detecting the expression of one or more biomarkers associated with innate immune cells and/or adaptive immune cells associated with each SQ subtype in a sample (*e.g.*, lung

cancer sample) obtained from a patient. In one embodiment, immune cell activation associated with a SQ subtype is determined by monitoring the immune cell signatures of Bindea et al (Immunity 2013; 39(4); 782-795), the contents of which are herein incorporated by reference in its entirety. In one embodiment, the method further comprises measuring single gene immune biomarkers, such as, for example, CTLA4, PDCD1 and CD274 (PD-L1), PDCDLG2(PD-L2) and/or IFN gene signatures. The presence or a detectable level of immune activation (Innate and/or Adaptive) associated with a SQ subtype can indicate or predict that a patient with said SQ subtype may be amendable to immunotherapy. The immunotherapy can be treatment with a checkpoint inhibitor as provided herein. In one embodiment, the primitive subtype of SQ has immune expression. In one embodiment, a method is provided herein for detecting the expression of at least one classifier biomarker provided herein in a sample (e.g., lung cancer SQ sample) obtained from a patient further comprises administering an immunotherapeutic agent following detection of immune activation as provided herein in said sample.

**[00172]** In one embodiment, the method comprises determining a subtype of a lung cancer SQ sample and subsequently determining a level of immune cell activation of said sub-type. In one embodiment, the subtype is determined by determining the expression levels of one or more classifier biomarkers using sequencing (e.g., RNASeq), amplification (e.g., qRT-PCR) or hybridization assays (e.g., microarray analysis) as described herein. The one or more biomarkers can be selected from a publically available database (e.g., TCGA lung SQ RNASeq gene expression datasets or any other publically available SQ gene expression datasets provided herein). In some embodiments, the biomarkers of **Table 1** can be used to specifically determine the subtype of an SQ lung sample obtained from a patient. In one embodiment, the level of immune cell activation is determined by measuring gene expression signatures of immunomarkers. The immunomarkers can be measured in the same and/or different sample used to subtype the lung cancer sample as described herein. The immunomarkers that can be measured can comprise, consist of, or consistently essentially of innate immune cell (IIC) and/or adaptive immune cell (AIC) gene signatures, interferon (IFN) gene signatures, individual immunomarkers, major histocompatibility complex class II (MHC class II) genes or a combination thereof. The gene expression signatures for both IICs and AICs can be any known gene signatures for said cell types known in the art. For example, the immune gene signatures can be those from Bindea et al. (Immunity 2013; 39(4); 782-795). In one embodiment, the immunomarkers for use in the methods provided herein

are selected from **Table 4A** and/or **Table 4B**. The individual immunomarkers can be CTLA4, PDCD1 and CD274 (PD-L1). In one embodiment, the individual immunomarkers for use in the methods provided herein are selected from **Table 5**. The immunomarkers can be one or more interferon (INF) genes. In one embodiment, the immunomarkers for use in the methods provided herein are selected from **Table 6**. The immunomarkers can be one or more MHCII genes. In one embodiment, the immunomarkers for use in the methods provided herein are selected from **Table 7**. In yet another embodiment, the immunomarkers for use in the methods provided herein are selected from **Tables 4A, 4B, 5, 6, 7**, or a combination thereof.

**Table 4A.** Adaptive immune cell (AIC) gene signature immunomarkers for use in the methods provided herein.

| Cell Type                                      |                                                                   |                                                    |                                                                        |                                                                             |                                                                                     |                                                                                  |  |
|------------------------------------------------|-------------------------------------------------------------------|----------------------------------------------------|------------------------------------------------------------------------|-----------------------------------------------------------------------------|-------------------------------------------------------------------------------------|----------------------------------------------------------------------------------|--|
|                                                | B cells                                                           | T cells                                            | T helper cells                                                         | Tcm                                                                         | Tem                                                                                 | Th1 cells                                                                        |  |
| Human Gene (Gene Name; GenBank Accession No.*) | ABCB4 (ATP binding cassette subfamily B member 4; NM_000443)      | BCL11B (B-cell lymphoma/leukaemia 11B; AJ404614.1) | ANP32B (acidic nuclear phosphoprotein 32 family member B; NM_006401.2) | AQP3 (aquaporine 3; NM_004925.4)                                            | AKT3 (AKT serine/threonine kinase 3; NM_005465.4)                                   | APBB2 (amyloid beta precursor protein binding family B member 2; NM_001166054.1) |  |
|                                                | BACH2 (BTB domain and CNC homolog 2; NM_021813.3)                 | CD2 (CD2 molecule; NM_001328609.1)                 | ASF1A (anti-silencing function 1A histone chaperone; NM_014034.2)      | ATF7IP (activating transcription factor 7 interacting protein; NM_181352.1) | C7orf54 (staphylococcal nuclease and tudor domain containing 1 (SND1); NG_051199.1) | APOD (apolipoprotein D; NM_001647.3)                                             |  |
|                                                | BCL11A (B-cell CLL/lymphoma 11A; NM_022893.3)                     | CD28 (CD28 molecule; NM_001243078.1)               | ATF2 (activating transcription factor 2; NM_001256093.1)               | ATM (ATM serine/threonine kinase; NM_000051.3)                              | CCR2 (C-C motif chemokine receptor 2; NM_001123396.1)                               | ATP9A (ATPase phospholipid transporting 9A; NM_006045.2)                         |  |
|                                                | BLK (BLK proto-oncogene, Src family tyrosine kinase; NM_001715.2) | CD3D (CD3d molecule; NM_000732.4)                  | BATF (basic leucine zipper ATF-like transcription factor; NM_006399.3) | CASP8 (caspase 8; NM_001228.4)                                              | DDX17 (DEAD-box helicase 17; NM_006386.4)                                           | BST2 (bone marrow stromal cell antigen 2; NM_004335.3)                           |  |
|                                                | BLNK (B-cell linker; NM_013314.3)                                 | CD3E (CD3e molecule; NM_000733.3)                  | C13orf34 (aurora borealis; EU834129.1)                                 | CDC14A (cell division cycle 14A; NM_003672.3)                               | EWSR1 (EWS RNA binding protein 1; NM_013986.3)                                      | BTG3 (BTG anti-proliferation factor 3; NM_001130914.1)                           |  |
|                                                | CCR9 (C-C motif chemokine receptor 9; NM_031200.2)                | CD3G (CD3g molecule; NM_000073.2)                  | CD28 (CD28 molecule; NM_006139.3)                                      | CEP68 (centrosomal protein 68; NM_015147.2)                                 | FLI1 (Fli-1 proto-oncogene, ETS transcription factor; NM_001130914.1)               | CCL4 (C-C motif chemokine ligand 4; NM_002984.3)                                 |  |
|                                                |                                                                   |                                                    |                                                                        |                                                                             |                                                                                     |                                                                                  |  |
|                                                |                                                                   |                                                    |                                                                        |                                                                             |                                                                                     |                                                                                  |  |

|                                                                    |                                                                      |                                                                     |                                                             |                                                                                   |                                                                                          |              |  |
|--------------------------------------------------------------------|----------------------------------------------------------------------|---------------------------------------------------------------------|-------------------------------------------------------------|-----------------------------------------------------------------------------------|------------------------------------------------------------------------------------------|--------------|--|
|                                                                    |                                                                      |                                                                     |                                                             |                                                                                   |                                                                                          | NM_002017.4) |  |
| CD19 (CD19 molecule; NM_001178098.1)                               | CD6 (CD6 molecule; NM_006725.4)                                      | DDX50 (DEAD-box helicase 50; NM_024045.1)                           | CG030 (BRCA2 region, mRNA sequence CG030; U50531.1)         | GDPD5 (glycerophosphodie ster phosphodiesterase domain containing 5; NM_030792.6) | CD38 (CD38 molecule; NM_001775.3)                                                        |              |  |
| CD72 (CD72 molecule; NM_001782.2)                                  | CD96 (CD96 molecule; NM_198196.2)                                    | FAM111A (family with sequence similarity 111 member A; NM_022074.3) | CLUAP1 (clusterin associated protein 1; NM_015041.2)        | LTK (leukocyte receptor tyrosine kinase; NM_002344.5)                             | CD70 (CD70 molecule; NM_001252.4)                                                        |              |  |
| COCH (cochlin; NM_001135058.1)                                     | GIMAP5 (GTPase, IMAP family member 5; NM_018384.4)                   | FRYL (FRY like transcription coactivator; NM_015030.1)              | CREBZF (CREB/ATF bZIP transcription factor; NM_001039618.2) | MEFV (Mediterranean fever; NM_000243.2)                                           | CMAH (cytidine monophospho-N-acetylneuraminic acid hydroxylase, pseudogene; NR_002174.2) |              |  |
| CR2 (complement C3d receptor 2; NM_001006658.2)                    | ITM2A (integral membrane protein 2A; NM_004867.4)                    | FUSIP1 (serine and arginine rich splicing factor 10; NM_006625.5)   | CYLD (CYLD lysine 63 deubiquitinase; NM_015247.2)           | NFATC4 (nuclear factor of activated T-cells 4; NM_001136022.2)                    | CSF2 (colony stimulating factor 2; NM_000758.3)                                          |              |  |
| DTNB (dystrobrevin beta; NM_021907.4)                              | LCK (LCK proto-oncogene, Src family tyrosine kinase; NM_001042771.2) | GOLGA8A (golgin A8 family member A; NM_181077.3)                    | CYorf15B (taxilin gamma pseudogene, Y-linked; NR_045128.1)  | PRKY (protein kinase, Y-linked, pseudogene; NR_028062.1)                          | CTLA4 (cytotoxic T-lymphocyte associated protein 4; NM_005214.4)                         |              |  |
| FAM30A (family with sequence similarity 30, member A; NR_026800.2) | NCALD (neurocalcin delta; NM_001040624.1)                            | ICOS (inducible T-cell costimulator; NM_012092.3)                   | DOCK9 (dedicator of cytokinesis 9; NM_015296.2)             | TBC1D5 (TBC1 domain family member 5; NM_001134381.1)                              | DGKI (diacylglycerol kinase iota; NM_004717.3)                                           |              |  |
| FCRL2 (Fc receptor like 2;                                         | PRKCQ (protein kinase C theta;                                       | ITM2A (integral membrane protein                                    | FOXP1 (forkhead box P1; NM_032682.5)                        | TBCD (tubulin folding cofactor D;                                                 | DOK5 (docking protein 5;                                                                 |              |  |

|                                                                                |                                                                         |                                                                              |                                                                                                     |                                                             |                                                     |              |
|--------------------------------------------------------------------------------|-------------------------------------------------------------------------|------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------|-------------------------------------------------------------|-----------------------------------------------------|--------------|
| NM_030764.3)                                                                   | NM_006257.4)                                                            | 2A; NM_004867.4)                                                             |                                                                                                     |                                                             | NM_005993.4)                                        | NM_018431.4) |
| GLDC (glycine decarboxylase; NM_000170.2)                                      | SH2D1A (SH2 domain containing 1A; NM_002351.4)                          | LRBA (LPS responsive beige-like anchor protein; NM_001199282.2)              | FYB (FYN binding protein; NM_001465.4)                                                              | TRA (T cell receptor alpha delta locus; NG_001332.3)        | DPP4 (dipeptidyl peptidase 4; NM_001935.3)          |              |
| GNG7 (G protein subunit gamma 7; NM_052847.2)                                  | SKAP1 (src kinase associated phosphoprotein 1; NM_001075099.1)          | NAP1L4 (nucleosome assembly protein 1 like 4; NM_005969.3)                   | HNRPH1 (heterogeneous nuclear ribonucleoprotein H1 (H); NM_001257293.1)                             | VIL2 (ezrin; NM_003379.4)                                   | DUSP5 (dual specificity phosphatase 5; NM_004419.3) |              |
| HLA-DOB (major histocompatibility complex, class II, DO beta; NM_002120.3)     | TRA (T cell receptor alpha delta locus; NG_001332.3)                    | NUP107 (nucleoporin 107; NM_020401.3)                                        | INPP4B (inositol polyphosphate-4-phosphatase type II B; NM_003866.3)                                | EGFL6 (EGF like domain multiple 6; NM_015507.3)             |                                                     |              |
| HLA-DQA1 (major histocompatibility complex, class II, DQ alpha 1; NM_002122.3) | TRAC (nuclear receptor corepressor 2; NM_006312.5)                      | PHF10 (PHD finger protein 10; NM_018288.3)                                   | KLF12 (Kruppel like factor 12; NM_007249.4)                                                         | GGT1 (gamma-glutamyltransferase 1; NM_013421.2)             |                                                     |              |
| IGHA1 (immunoglobulin heavy locus; NG_001019.6)                                | TRAT1 (T cell receptor associated transmembrane adaptor 1; NM_016388.3) | PPP2R5C (protein phosphatase 2 regulatory subunit B', gamma; NM_001161725.1) | LOC202134 (family with sequence similarity 153 member B; NM_001265615.1)                            | HBEGF (heparin binding EGF like growth factor; NM_001945.2) |                                                     |              |
| IGHG1 (immunoglobulin heavy locus; NG_001019.6)                                | TRBC1 (T cell receptor beta locus; NG_001333.2)                         | RPA1 (replication protein A1; NM_002945.3)                                   | MAP3K1 (mitogen-activated protein kinase kinase kinase 1, E3 ubiquitin protein ligase; NM_005921.1) | IFNG (interferon gamma; NM_000619.2)                        |                                                     |              |

|  |                                                                                                        |  |                                                                     |                                                                                         |  |                                                                                  |
|--|--------------------------------------------------------------------------------------------------------|--|---------------------------------------------------------------------|-----------------------------------------------------------------------------------------|--|----------------------------------------------------------------------------------|
|  | IGHM<br>(immunoglobulin heavy locus; NG_001019.6)                                                      |  | SEC24C (SEC24 homolog C, COPII coat complex component; NM_004922.3) | MLL (lysine (K)-specific methyltransferase 2A; NM_005933.3)                             |  | IL12RB2 (interleukin 12 receptor subunit beta 2; NM_001319233.1)                 |
|  | IGKC<br>(immunoglobulin kappa locus, proximal V-cluster and J-C cluster; NG_000834.1)                  |  | SLC25A12 (solute carrier family 25 member 12; NM_003705.4)          | NEFL (neurofilament, light polypeptide; NM_006158.4)                                    |  | IL22 (interleukin 22; NM_020525.4)                                               |
|  | IGL<br>(immunoglobulin lambda locus; NG_000002.1)                                                      |  | TRA (T cell receptor alpha delta locus; NG_001332.3)                | NFATC3 (nuclear factor of activated T-cells 3; NM_173165.2)                             |  | LRP8 (LDL receptor related protein 8; NM_017522.4)                               |
|  | KIAA0125 (family with sequence similarity 30, member A; NR_026800.2)                                   |  | UBE2L3 (ubiquitin conjugating enzyme E2 L3; NM_003347.3)            | PCM1 (pericentriolar material 1; NM_001315507.1)                                        |  | LRRN3 (leucine rich repeat neuronal 3; NM_018334.4)                              |
|  | MEF2C (myocyte enhancer factor 2C; NM_001308002.1)                                                     |  | YME1L1 (YME1 like 1 ATPase; NM_001253866.1)                         | PCNX (pecanex homolog 1; NM_014982.2)                                                   |  | LTA (lymphotoxin alpha; NM_000595.3)                                             |
|  | MICAL3<br>(microtubule associated monooxygenase, calponin and LIM domain containing 3; NM_001136004.3) |  |                                                                     | PDXDC2 (pyridoxal dependent decarboxylase domain containing 2, pseudogene; NR_003610.1) |  | SGCB (sarcoglycan, beta (43kDa dystrophin-associated glycoprotein); NM_000232.4) |
|  | MS4A1 (membrane spanning 4-domains                                                                     |  |                                                                     | PHC3 (polyhomeotic homolog 3;                                                           |  | SYNGR3 (synaptogyrin 3;                                                          |
|  |                                                                                                        |  |                                                                     |                                                                                         |  |                                                                                  |
|  |                                                                                                        |  |                                                                     |                                                                                         |  |                                                                                  |
|  |                                                                                                        |  |                                                                     |                                                                                         |  |                                                                                  |

|  |                                                                              |  |  |  |                                                                             |  |                                                                |
|--|------------------------------------------------------------------------------|--|--|--|-----------------------------------------------------------------------------|--|----------------------------------------------------------------|
|  | A1; NM_021950.3)                                                             |  |  |  | NM_001308116.1)                                                             |  | NM_004209.5)                                                   |
|  | OSBPL10 (oxysterol binding protein like 10; NM_017784.4)                     |  |  |  | POLR2J2 (RNA polymerase II subunit J2; NM_032959.5)                         |  | ZBTB32 (zinc finger and BTB domain containing 32; NM_014383.2) |
|  | PNOC (prepronociceptin; NM_001284244.1)                                      |  |  |  | PSPC1 (paraspeckle component 1; NM_001042414.2)                             |  |                                                                |
|  | QRSL1 (glutaminyl-tRNA synthase (glutamine-hydrolyzing)-like 1; NM_018292.4) |  |  |  | REPS1 (RALBP1 associated Eps domain containing 1; NM_001128617.2)           |  |                                                                |
|  | SCN3A (sodium voltage-gated channel alpha subunit 3; NM_001081677.1)         |  |  |  | RP11-74E24.2 (zinc finger CCCH-type domain-containing-like; NM_001271675.1) |  |                                                                |
|  | SLC15A2 (solute carrier family 15 member 2; XM_017007074.1)                  |  |  |  | RPP38 (ribonuclease P/MRP subunit p38; NM_001265601.1)                      |  |                                                                |
|  | SPIB (Spi-B transcription factor; NM_001244000.1)                            |  |  |  | SLC7A6 (solute carrier family 7 member 6; NM_003983.5)                      |  |                                                                |
|  | TCL1A (T-cell leukemia/lymphoma 1A; NM_001098725.1)                          |  |  |  | SNRPN (small nuclear ribonucleoprotein polypeptide N; NM_022807.3)          |  |                                                                |
|  | TNFRSF17 (TNF receptor superfamily                                           |  |  |  | ST3GAL1 (ST3 beta-galactoside alpha-2,3-sialyltransferase                   |  |                                                                |
|  |                                                                              |  |  |  |                                                                             |  |                                                                |
|  |                                                                              |  |  |  |                                                                             |  |                                                                |
|  |                                                                              |  |  |  |                                                                             |  |                                                                |
|  |                                                                              |  |  |  |                                                                             |  |                                                                |

|  |                            |  |  |  |                                                                                                 |  |  |
|--|----------------------------|--|--|--|-------------------------------------------------------------------------------------------------|--|--|
|  | member 17;<br>NM_001192.2) |  |  |  | 1; NM_173344.2)                                                                                 |  |  |
|  |                            |  |  |  | STX16 (syntaxin 16;<br>NM_001204868.1)                                                          |  |  |
|  |                            |  |  |  | TIMM8A (translocase<br>of inner<br>mitochondrial<br>membrane 8<br>homolog A;<br>NM_001145951.1) |  |  |
|  |                            |  |  |  | TRAF3IP3 (TRAF3<br>interacting protein 3;<br>NM_001320144.1)                                    |  |  |
|  |                            |  |  |  | TXK (TXK tyrosine<br>kinase;<br>NM_003328.2)                                                    |  |  |
|  |                            |  |  |  | USP9Y (ubiquitin<br>specific peptidase 9,<br>Y-linked;<br>NG_008311.1)                          |  |  |

\*Each GenBank Accession Number is a representative or exemplary GenBank Accession Number for the listed gene and is herein incorporated by reference in its entirety for all purposes. Further, each listed representative or exemplary accession number should not be construed to limit the claims to the specific accession number.

| Table 4A continued     |                                             |                                                        |                                      |                                      |                                                        |                                              |                                                               |
|------------------------|---------------------------------------------|--------------------------------------------------------|--------------------------------------|--------------------------------------|--------------------------------------------------------|----------------------------------------------|---------------------------------------------------------------|
| Cell Type              |                                             |                                                        |                                      |                                      |                                                        |                                              |                                                               |
|                        | Th2 cells                                   | TFH                                                    | Th17 cells                           | TReg                                 | CD8 T cells                                            | Tgd                                          | Cytotoxic cells                                               |
| Human Gene (Gene Name; | ADCY1 (adenylate cyclase 1; NM_001281768.1) | B3GAT1 (beta-1,3-glucuronyltransferase 1; NM_018644.3) | IL17A (interleukin 17A; NM_002190.2) | FOXP3 (forkhead box P3; NM_014009.3) | ABT1 (activator of basal transcription 1; NM_013375.3) | C1orf61 (chromosome 1 open reading frame 61; | APBA2 (amyloid beta precursor protein binding family A member |

|                                |                                                                                         |                                                                                                 |                                                                   |  |                                                                                              |                                                                                  |                                                              |
|--------------------------------|-----------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------|-------------------------------------------------------------------|--|----------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------|--------------------------------------------------------------|
| GenBank<br>Accessio<br>n No.*) | AHI1 (Abelson<br>helper<br>integration site 1;<br>NM_001134831.<br>1)                   | BLR1 (c-x-c<br>chemokine receptor<br>type 5; EF444957.1)                                        | IL17RA<br>(interleukin 17<br>receptor A;<br>NM_014339.6)          |  | AES (amino-<br>terminal enhancer<br>of split;<br>NM_198969.1)                                | CD160 (CD160<br>molecule;<br>NM_007053.3)                                        | 2; NM_005503.3)                                              |
|                                | AI582773<br>(tn17d08.x1<br>NCI_CGAP_Brn25<br>Homo sapiens<br>cDNA clone;<br>AI582773.1) | C18orf1 (low density<br>lipoprotein receptor<br>class A domain<br>containing 4;<br>NM_181481.4) | RORC (RAR<br>related orphan<br>receptor C;<br>NM_001001523.<br>1) |  | APBA2 (amyloid<br>beta precursor<br>protein binding<br>family A member 2;<br>NM_001130414.1) | FEZ1<br>(Fasciculation<br>And Elongation<br>Protein Zeta 1;<br>AF123659.1)       | CTSW (cathepsin<br>W;<br>NM_001335.3)                        |
|                                | ANK1 (ankyrin 1;<br>NM_020476.2)                                                        | CDK5R1 (cyclin<br>dependent kinase 5<br>regulatory subunit<br>1; NM_003885.2)                   |                                                                   |  | ARHGAP8 (Rho<br>GTPase activating<br>protein 8;<br>NM_001198726.1)                           | TARP (TCR<br>gamma alternate<br>reading frame<br>protein;<br>NM_001003806.<br>1) | DUSP2 (dual<br>specificity<br>phosphatase 2;<br>NM_004418.3) |
|                                | BIRC5<br>(baculoviral IAP<br>repeat containing<br>5;<br>NM_001012271.<br>1)             | CHGB<br>(chromogranin B;<br>NM_001819.2)                                                        |                                                                   |  | C12orf47<br>(MAPKAPK5<br>antisense RNA 1;<br>NR_015404.1)                                    | TRD (T cell<br>receptor alpha<br>delta locus;<br>NG_001332.3)                    | GNLY (granulysin;<br>NM_012483.3)                            |
|                                | CDC25C (cell<br>division cycle<br>25C;<br>NM_001318098.<br>1)                           | CHI3L2 (chitinase 3<br>like 2;<br>NM_001025199.1)                                               |                                                                   |  | C19orf6<br>(transmembrane<br>protein 259;<br>NM_001033026.1)                                 | TRGV9 (T cell<br>receptor gamma<br>V region 9;<br>X69385.1)                      | GZMA (granzyme<br>A; NM_006144.3)                            |
|                                | CDC7 (cell<br>division cycle 7;<br>NM_001134420.<br>1)                                  | CXCL13 (C-X-C motif<br>chemokine ligand<br>13; NM_006419.2)                                     |                                                                   |  | C4orf15 (HAUS<br>augmin like<br>complex subunit 3;<br>NM_001303143.1)                        |                                                                                  | GZMH (granzyme<br>H;<br>NM_001270781.<br>1)                  |

|                                                                        |                                                                                                   |  |  |                                                                                       |                                                                          |
|------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------|--|--|---------------------------------------------------------------------------------------|--------------------------------------------------------------------------|
| CENPF<br>(centromere<br>protein F;<br>NM_016343.3)                     | HEY1 (hes related<br>family bHLH<br>transcription factor<br>with YRPW motif 1;<br>NM_001282851.1) |  |  | CAMLG (calcium<br>modulating ligand;<br>NM_001745.3)                                  | KLRB1 (killer cell<br>lectin like<br>receptor B1;<br>NM_002258.2)        |
| CXCR6 (killer cell<br>lectin like<br>receptor B1;<br>NM_002258.2)      | HIST1H4K (histone<br>cluster 1 H4 family<br>member k;<br>NM_003541.2)                             |  |  | CD8A (CD8a<br>molecule;<br>NM_001768.6)                                               | KLRD1 (killer cell<br>lectin like<br>receptor D1;<br>NM_001114396.<br>1) |
| DHFR<br>(dihydrofolate<br>reductase;<br>NM_001290354.<br>1)            | ICA1 (islet cell<br>autoantigen 1;<br>NM_001136020.2)                                             |  |  | CD8B (CD8b<br>molecule;<br>NM_001178100.1)                                            | KLRF1 (killer cell<br>lectin like<br>receptor F1;<br>NM_001291822.<br>1) |
| EVI5 (ecotropic<br>viral integration<br>site 5;<br>NM_001308248.<br>1) | KCNK5 (potassium<br>two pore domain<br>channel subfamily K<br>member 5;<br>NM_003740.3)           |  |  | CDKN2AIP<br>(CDKN2A<br>interacting protein;<br>NM_001317343.1)                        | KLRK1 (killer cell<br>lectin like<br>receptor K1;<br>NM_007360.3)        |
| GATA3 (GATA<br>binding protein 3;<br>NM_001002295.<br>1)               | KIAA1324<br>(KIAA1324;<br>NM_001284353.1)                                                         |  |  | DNAJB1 (DnaJ heat<br>shock protein<br>family (Hsp40)<br>member B1;<br>NM_001313964.1) | NKG7 (natural<br>killer cell granule<br>protein 7;<br>NM_005601.3)       |
| GSTA4<br>(glutathione S-<br>transferase alpha<br>4; NM_001512.3)       | MAF (MAF bZIP<br>transcription factor;<br>NM_001031804.2)                                         |  |  | FLT3LG (fms<br>related tyrosine<br>kinase 3 ligand;<br>NM_001278638.1)                | RORA (RAR<br>related orphan<br>receptor A;<br>NM_134262.2)               |
| HELLS (helicase,<br>lymphoid-<br>specific;<br>NM_001289074.            | MAGEH1 (MAGE<br>family member H1;<br>NM_014061.4)                                                 |  |  | GADD45A (growth<br>arrest and DNA<br>damage inducible<br>alpha;                       | RUNX3 (runt<br>related<br>transcription<br>factor 3;                     |

|                                                                                                   |                                                                      |  |  |  |  |                                                                            |  |                                                                                                                 |
|---------------------------------------------------------------------------------------------------|----------------------------------------------------------------------|--|--|--|--|----------------------------------------------------------------------------|--|-----------------------------------------------------------------------------------------------------------------|
| 1)                                                                                                |                                                                      |  |  |  |  | NM_001199742.1)                                                            |  | NM_004350.2)                                                                                                    |
| IL26 (interleukin 26; NM_018402.1)                                                                | MKL2 (MKL1/myocardin like 2; NM_014048.4)                            |  |  |  |  | GZMM (granzyme M; NM_001258351.1)                                          |  | SIGIRR (single Ig and TIR domain containing; NM_001135054.1)                                                    |
| LAIR2 (leukocyte associated immunoglobulin like receptor 2; NM_021270.4)                          | MYO6 (myosin VI; NM_001300899.1)                                     |  |  |  |  | KLF9 (Kruppel like factor 9; NM_001206.2)                                  |  | WHDC1L1 (WAS protein homolog associated with actin, golgi membranes and microtubules pseudogene 3; NR_003521.1) |
| LIMA1 (LIM domain and actin binding 1; NM_001243775.1)                                            | MYO7A (myosin VIIA; NM_001127179.2)                                  |  |  |  |  | LEPROTL1 (leptin receptor overlapping transcript-like 1; NM_001128208.1)   |  | ZBTB16 (zinc finger and BTB domain containing 16; NM_001018011.1)                                               |
| MB (myoglobin; NM_203377.1)                                                                       | PASK (PAS domain containing serine/threonine kinase; NM_001252119.1) |  |  |  |  | LIME1 (Lck interacting transmembrane adaptor 1; NM_017806.3)               |  |                                                                                                                 |
| MICAL2 (microtubule associated monooxygenase, calponin and LIM domain containing 2; NM_001282663. | PDCD1 (programmed cell death 1; NM_005018.2)                         |  |  |  |  | MYST3 (MYST histone acetyltransferase (monocytic leukemia) 3; NM_006766.4) |  |                                                                                                                 |



|                                                                                      |                                                                                                          |  |  |  |                                                                       |  |  |
|--------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------|--|--|--|-----------------------------------------------------------------------|--|--|
| n D1 polypeptide;<br>NM_001291916.<br>1)                                             |                                                                                                          |  |  |  | NM_001031684.2)                                                       |  |  |
| WDHD1 (WD<br>repeat and HMG-<br>box DNA binding<br>protein 1;<br>NM_001008396.<br>2) | ST8SIA1 (ST8 alpha-<br>N-acetyl-<br>neuraminide alpha-<br>2,8-sialyltransferase<br>1;<br>NM_001304450.1) |  |  |  | SLC16A7 (solute<br>carrier family 16<br>member 7;<br>NM_001270622.1)  |  |  |
|                                                                                      | STK39<br>(serine/threonine<br>kinase 39;<br>NM_013233.2)                                                 |  |  |  | TBCC (tubulin<br>folding cofactor C;<br>NM_003192.2)                  |  |  |
|                                                                                      | THADA (THADA,<br>armadillo repeat<br>containing;<br>NM_001271644.1)                                      |  |  |  | THUMPD1 (THUMP<br>domain containing<br>1; NM_017736.4)                |  |  |
|                                                                                      | TOX (thymocyte<br>selection associated<br>high mobility group<br>box; NM_014729.2)                       |  |  |  | TMC6<br>(transmembrane<br>channel like 6;<br>NM_001321185.1)          |  |  |
|                                                                                      | TSHR (thyroid<br>stimulating<br>hormone receptor;<br>NM_000369.2)                                        |  |  |  | TSC2D3 (TSC22<br>domain family<br>member 3;<br>NM_001318470.1)        |  |  |
|                                                                                      | ZNF764 (zinc finger<br>protein 764;<br>NM_001172679.1)                                                   |  |  |  | VAMP2 (vesicle<br>associated<br>membrane protein<br>2; NM_014232.2)   |  |  |
|                                                                                      |                                                                                                          |  |  |  | ZEB1 (zinc finger E-<br>box binding<br>homeobox 1;<br>NM_001128128.2) |  |  |

|  |  |  |  |  |  |                                                         |  |  |  |
|--|--|--|--|--|--|---------------------------------------------------------|--|--|--|
|  |  |  |  |  |  | ZFP36L2 (ZFP36 ring finger protein like 2; NM_006887.4) |  |  |  |
|  |  |  |  |  |  | ZNF22 (zinc finger protein 22; NM_006963.4)             |  |  |  |
|  |  |  |  |  |  | ZNF609 (zinc finger protein 609; NM_015042.1)           |  |  |  |
|  |  |  |  |  |  | ZNF91 (zinc finger protein 91; NM_001300951.1)          |  |  |  |

\*Each GenBank Accession Number is a representative or exemplary GenBank Accession Number for the listed gene and is herein incorporated by reference in its entirety for all purposes. Further, each listed representative or exemplary accession number should not be construed to limit the claims to the specific accession number.

[00174]                   **Table 4B.** Innate immune cell (IIC) gene signature immunomarkers for use in the methods provided herein.

| Cell Type                                      |                                                          |                                                           |                                                                   |                                                    |                                                                                                   |  |
|------------------------------------------------|----------------------------------------------------------|-----------------------------------------------------------|-------------------------------------------------------------------|----------------------------------------------------|---------------------------------------------------------------------------------------------------|--|
|                                                | NK cells                                                 | NK CD56dim cells                                          | NK CD56bright cells                                               | DC                                                 | iDC                                                                                               |  |
| Human Gene (Gene Name; GenBank Accession No.*) | ADARB1 (adenosine deaminase, RNA specific B1; NM_001112) | EDG8 (sphingosine-1-phosphate receptor 5; NM_001166215.1) | BG255923 (lysophosphatidylcholine acyltransferase 4; NM_153613.2) | CCL13 (C-C motif chemokine ligand 13; NM_005408.2) | ABCG2 (ATP-binding cassette, sub-family G (WHITE), member 2 (Junior blood group); NM_001257386.1) |  |
|                                                | AF107846 (neuroendocrine-specific Golgi protein p55;     | FLJ20699 (cDNA FLJ20699 fis, clone KAI2372; AK000706.1)   | DUSP4 (dual specificity phosphatase 4; NM_057158.3)               | CCL17 (C-C motif chemokine ligand 17; NM_002987.2) | BLVRB (biliverdin reductase B; NM_000713.2)                                                       |  |

|                                                                    |                                                                                                                 |                                                                           |                                                                  |                                                                 |  |  |
|--------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------|------------------------------------------------------------------|-----------------------------------------------------------------|--|--|
| AF107846.1)                                                        |                                                                                                                 |                                                                           |                                                                  |                                                                 |  |  |
| AL080130 (cDNA DKFZp434E033 (from clone DKFZp434E033); AL080130.1) | GTF3C1 (general transcription factor IIIC subunit 1; NM_001286242.1)                                            | FOXJ1 (forkhead box J1; NM_001454.3)                                      | CCL22 (C-C motif chemokine ligand 22; NM_002990.4)               | CARD9 (caspase recruitment domain family member 9; NM_052814.3) |  |  |
| ALDH1B1 (aldehyde dehydrogenase 1 family member B1; NM_000692.4)   | GZMB (granzyme B; NM_004131.4)                                                                                  | MADD (MAP kinase activating death domain; NM_001135944.1)                 | CD209 (CD209 molecule; NM_001144899.1)                           | CD1A (CD1a molecule; NM_001763.2)                               |  |  |
| ARL6IP2 (atlastin GTPase 2; NM_001330461.1)                        | IL21R (interleukin 21 receptor; NM_181079.4)                                                                    | MPPED1 (metallophosphoesterase domain containing 1, mRNA; NM_001044370.1) | HSD11B1 (hydroxysteroid 11-beta dehydrogenase 1; NM_001206741.1) | CD1B (CD1b molecule; NM_001764.2)                               |  |  |
| BCL2 (apoptosis regulator (BCL2); NM_000633.2)                     | KIR2DL3 (killer cell immunoglobulin like receptor, two Ig domains and long cytoplasmic tail 3; NM_015868.2)     | MUC3B (mucin 3B cell surface associated; JQ511939.1)                      | NPR1 (natriuretic peptide receptor 1; NM_000906.3)               | CD1C (CD1c molecule; NM_001765.2)                               |  |  |
| CDC5L (cell division cycle 5 like; NM_001253.3)                    | KIR2DS1 (killer cell immunoglobulin like receptor, two Ig domains and short cytoplasmic tail 1; NM_014512.1)    | NIBP (NIK and IKKbeta-binding protein; AY630619.1)                        | PPFIBP2 (PPFIA binding protein 2; XR_930917.2)                   | CD1E (CD1e molecule; NM_001185115.1)                            |  |  |
| FGF18 (fibroblast growth factor 18; NM_003862.2)                   | KIR2DS2 (killer cell immunoglobulin like receptor, two Ig domains and short cytoplasmic tail 2; NM_001291700.1) | PLA2G6 (phospholipase A2 group VI; NM_001004426.1)                        |                                                                  | CH25H (cholesterol 25-hydroxylase; NM_003956.3)                 |  |  |
| FUT5 (fucosyltransferase 5; NM_002034.2)                           | KIR2DS5 (killer cell immunoglobulin like                                                                        | RRAD (Ras related glycolysis inhibitor and                                |                                                                  | CLEC10A (C-type lectin domain family 10 member                  |  |  |

|                                                                    |                                                                                                                   |                                                  |  |                                                                  |
|--------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------|--------------------------------------------------|--|------------------------------------------------------------------|
|                                                                    | receptor, two Ig domains and short cytoplasmic tail 5; NM_014513.2)                                               | calcium channel regulator; NM_001128850.1)       |  | A; NM_001330070.1)                                               |
| FZR1 (fizzy/cell division cycle 20 related 1; XM_005259573.4)      | KIR3DL1 (killer cell immunoglobulin like receptor, three Ig domains and long cytoplasmic tail 1; NM_013289.2)     | SEPT6 (septin 6; NM_145802.3)                    |  | CSF1R (colony stimulating factor 1 receptor; NM_001288705.1)     |
| GAGE2 (G antigen 2; NM_001127212.1)                                | KIR3DL2 (killer cell immunoglobulin like receptor, three Ig domains and long cytoplasmic tail 2; NM_006737.3)     | XCL1 (X-C motif chemokine ligand 1; NM_002995.2) |  | CTNS (cystinosis, lysosomal cystine transporter; NM_001031681.2) |
| IGFBP5 (insulin like growth factor binding protein 5; NM_000599.3) | KIR3DL3 (killer cell immunoglobulin like receptor, three Ig domains and long cytoplasmic tail 3; NM_153443.4)     |                                                  |  | F13A1 (factor XIII a subunit; AH002691.2)                        |
| LDB3 (LIM domain binding 3; NM_001171611.1)                        | KIR3DS1 (killer cell immunoglobulin like receptor, three Ig domains and short cytoplasmic tail 1; NM_001083539.2) |                                                  |  | FABP4 (fatty acid binding protein 4; NM_001442.2)                |
| LOC643313 (similar to hypothetical protein LOC284701; XM_933043.1) | SPON2 (spondin 2; NM_001199021.1)                                                                                 |                                                  |  | FZD2 (frizzled class receptor 2; NM_001466.3)                    |
| LOC730096 (hypothetical protein LOC730096;                         | TMEPAI (prostate transmembrane                                                                                    |                                                  |  | GSTT1 (glutathione S-transferase theta 1;                        |

|                                                                                             |                                           |  |  |  |                                                                            |
|---------------------------------------------------------------------------------------------|-------------------------------------------|--|--|--|----------------------------------------------------------------------------|
| NC_000022.9)                                                                                | protein, androgen induced 1; NM_199169.2) |  |  |  | NM_001293814.1)                                                            |
| MAPRE3 (microtubule associated protein RP/EB family member 3; NM_001303050.1)               |                                           |  |  |  | GUCA1A (guanylate cyclase activator 1A; NM_001319062.1)                    |
| MCM3AP (minichromosome maintenance complex component 3 associated protein; NM_003906.4)     |                                           |  |  |  | HS3ST2 (heparan sulfate (glucosamine) 3-O-sulfotransferase 2; NM_006043.1) |
| MRC2 (mannose receptor C type 2; NM_006039.4)                                               |                                           |  |  |  | LMAN2L (lectin, mannose binding 2 like; NM_001322355.1)                    |
| NCR1 (natural cytotoxicity triggering receptor 1; NM_001242357.2)                           |                                           |  |  |  | MMP12 (matrix metalloproteinase 12; NM_002426.5)                           |
| NM_014114 (PRO0097 protein; NM_014114.1)                                                    |                                           |  |  |  | MS4A6A (membrane spanning 4-domains A6A; NM_001330275.1)                   |
| NM_014274 (transient receptor potential cation channel, subfamily V, member 6; NM_014274.3) |                                           |  |  |  | NM_021941 (chromosome 21 open reading frame 97; NM_021941.1)               |
| NM_017616 (KN motif and ankyrin repeat domains 2; NM_015493.6)                              |                                           |  |  |  | NUDT9 (nudix hydrolase 9; NM_001248011.1)                                  |
| PDLIM4 (PDZ and LIM domain 4; NM_003687.3)                                                  |                                           |  |  |  | PPARG (peroxisome proliferator activated receptor gamma; NM_005037.5)      |
| PRX (periaxin;                                                                              |                                           |  |  |  | PREP (prolyl                                                               |

|                                                                       |  |  |  |  |  |                                                                           |
|-----------------------------------------------------------------------|--|--|--|--|--|---------------------------------------------------------------------------|
| NM_020956.2)                                                          |  |  |  |  |  | endopeptidase;<br>NM_002726.4)                                            |
| PSMD4 (proteasome 26S subunit, non-ATPase 4; NM_001330692.1)          |  |  |  |  |  | RAP1GAP (RAP1 GTPase activating protein; NM_001330383.1)                  |
| RP5-886K2.1 (neuronal thread protein AD7c-NTP; AF010144.1)            |  |  |  |  |  | SLC26A6 (solute carrier family 26 member 6; NM_001281733.1)               |
| SLC30A5 (solute carrier family 30 member 5; NM_001251969.1)           |  |  |  |  |  | SLC7A8 (solute carrier family 7 member 8; NR_049767.1)                    |
| SMEK1 (protein phosphatase 4 regulatory subunit 3A; NM_001284280.1)   |  |  |  |  |  | SYT17 (synaptotagmin 17; NM_001330509.1)                                  |
| SPN (sialophorin; NM_003123.4)                                        |  |  |  |  |  | TACSTD2 (tumor-associated calcium signal transducer 2; NM_002353.2)       |
| TBXA2R (thromboxane A2 receptor; NM_001060.5)                         |  |  |  |  |  | TM7SF4 (dendrocyte expressed seven transmembrane protein; NM_001257317.1) |
| TCTN2 (tectonic family member 2; NM_001143850.2)                      |  |  |  |  |  | VASH1 (vasohibin 1; NM_014909.4)                                          |
| TINAGL1 (tubulointerstitial nephritis antigen like 1; NM_001204415.1) |  |  |  |  |  |                                                                           |
| XCL1 (X-C motif chemokine ligand 1; NM_002995.2)                      |  |  |  |  |  |                                                                           |
| XCL2 (X-C motif chemokine ligand 2; NM_003175.3)                      |  |  |  |  |  |                                                                           |

|  |                                                  |  |  |  |  |
|--|--------------------------------------------------|--|--|--|--|
|  | ZNF205 (zinc finger protein 205; NM_001278158.1) |  |  |  |  |
|  | ZNF528 (zinc finger protein 528; NM_032423.2)    |  |  |  |  |
|  | ZNF747 (zinc finger protein 747; NM_023931.3)    |  |  |  |  |

\*Each GenBank Accession Number is a representative or exemplary GenBank Accession Number for the listed gene and is herein incorporated by reference in its entirety for all purposes. Further, each listed representative or exemplary accession number should not be construed to limit the claims to the specific accession number.

| Table 4B continued                              |                                                              |                                                              |                                                      |                                                                  |                                                                                  |                                                                                     |
|-------------------------------------------------|--------------------------------------------------------------|--------------------------------------------------------------|------------------------------------------------------|------------------------------------------------------------------|----------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|
| Cell Type                                       |                                                              |                                                              |                                                      |                                                                  |                                                                                  |                                                                                     |
|                                                 | aDC                                                          | pDC                                                          | Eosinophils                                          | Macrophages                                                      | Mast cells                                                                       | Neutrophils                                                                         |
| Human Gene (Gene Name; GenBank Accession No. *) | CCL1 (Chemokine (C-C motif) ligand 1; NM_002981)             | IL3RA (interleukin 3 receptor subunit alpha; NM_001267713.1) | ABHD2 (abhydrolase domain containing 2; NM_007011.7) | APOE (apolipoprotein E; NM_001302691.1)                          | ABCC4 (ATP binding cassette subfamily C member 4; NM_001301829.1)                | ALPL (alkaline phosphatase, liver/bone/kidney; NM_001127501.3)                      |
|                                                 | EBI3 (Epstein-Barr virus induced 3; NM_005755.2)             |                                                              | ACACB (acetyl-CoA carboxylase beta; NM_001093.3)     | ATG7 (autophagy related 7; NM_001144912.1)                       | ADCYAP1 (adenylate cyclase activating polypeptide 1; NM_001117.4)                | BST1 (bone marrow stromal cell antigen 1; NM_004334.2)                              |
|                                                 | INDO (indoleamine-pyrrole 2,3 dioxygenase; AY221100.1)       |                                                              | C9orf156 (tRNA methyltransferase O; NM_001330725.1)  | BCAT1 (branched chain amino acid transaminase 1; NM_001178094.1) | CALB2 (calbindin 2; NM_001740.4)                                                 | CD93 (CD93 molecule; NM_012072.3)                                                   |
|                                                 | LAMP3 (lysosomal associated membrane protein 3; NM_014398.3) |                                                              | CAT (catalase; NM_001752.3)                          | CCL7 (C-C motif chemokine ligand 7; NM_006273.3)                 | CEACAM8 (carcinoembryonic antigen related cell adhesion molecule 8; NM_001816.3) | CEACAM3 (carcinoembryonic antigen related cell adhesion molecule 3; NM_001277163.2) |

|                                                       |                                                                                  |                                                                 |                                                                    |                                                                                  |
|-------------------------------------------------------|----------------------------------------------------------------------------------|-----------------------------------------------------------------|--------------------------------------------------------------------|----------------------------------------------------------------------------------|
| OAS3 (2'-5'-oligoadenylate synthetase 3; NM_006187.3) | CCR3 (C-C motif chemokine receptor 3; NM_178329.2)                               | CD163 (CD163 molecule; NM_203416.3)                             | CMA1 (chymase 1, mast cell; NM_001308083.1)                        | CREB5 (cAMP responsive element binding protein 5; NM_001011666.2)                |
|                                                       | CLC (Charcot-Leyden crystal galectin; NM_001828.5)                               | CD68 (CD68 molecule; NM_001040059.1)                            | CPA3 (carboxypeptidase A3; NM_001870.3)                            | CRISPLD2 (cysteine rich secretory protein LCCL domain containing 2; NM_031476.3) |
|                                                       | CYSLTR2 (cysteinyll leukotriene receptor 2; NM_001308471.1)                      | CD84 (CD84 molecule; NM_001184881.1)                            | CTSG (cathepsin G; NM_001911.2)                                    | CSF3R (colony stimulating factor 3 receptor; NM_172313.2)                        |
|                                                       | EMR1 (EGF-like module containing mucin-like hormone receptor-like 1; DQ217942.1) | CHI3L1 (chitinase 3 like 1; NM_001276.2)                        | ELA2 (neutrophil elastase; EU617980.1)                             | CYP4F3 (cytochrome P450 family 4 subfamily F member 3; NM_001199209.1)           |
|                                                       | EPN2 (epsin 2; NM_001102664.1)                                                   | CHIT1 (chitinase 1; NM_001270509.1)                             | GATA2 (GATA binding protein 2; NM_001145661.1)                     | DYSF (dysferlin; NM_001130455.1)                                                 |
|                                                       | GALC (galactosylceramidase; NM_000153.3)                                         | CLEC5A (C-type lectin domain family 5 member A; NM_001301167.1) | HDC (histidine decarboxylase; NM_002112.3)                         | FCAR (Fc fragment of IgA receptor; NM_133278.3)                                  |
|                                                       | GPR44 (orphan G protein-coupled receptor; AF118265.1)                            | COL8A2 (collagen type VIII alpha 2 chain; NM_001294347.1)       | HPGD (hydroxyprostaglandin dehydrogenase 15-(NAD); NM_001256307.1) | FCGR3B (Fc fragment of IgG receptor IIIb; NM_001271035.1)                        |
|                                                       | HES1 (hes family bHLH transcription factor 1; NM_005524.3)                       | COLEC12 (collectin subfamily member 12; NM_130386.2)            | KIT (KIT proto-oncogene receptor tyrosine kinase; NM_001115115.1)  | FLJ11151 (hypothetical protein FLJ11151; BC006289.2)                             |

|  |  |                                                                                           |                                                     |                                                                      |                                                                                |
|--|--|-------------------------------------------------------------------------------------------|-----------------------------------------------------|----------------------------------------------------------------------|--------------------------------------------------------------------------------|
|  |  |                                                                                           |                                                     | NM_000222.2)                                                         |                                                                                |
|  |  | HIST1H1C (histone cluster 1 H1 family member c; NM_005319.3)                              | CTSK (cathepsin K; NM_000396.3)                     | LOC339524 (long intergenic non-protein coding RNA 1140; NR_026985.1) | FPR1 (formyl peptide receptor 1; NM_001193306.1)                               |
|  |  | HRH4 (histamine receptor H4; NM_001143828.1)                                              | CXCL5 (C-X-C motif chemokine ligand 5; NM_002994.4) | LOH11CR2A (BCSC-1 isoform; AY366508.1)                               | FPRL1 (formyl peptide receptor-like receptor; M84562.1)                        |
|  |  | IGSF2 (immunoglobulin superfamily, member 2; BC130327.1)                                  | CYBB (cytochrome b-245 beta chain; NM_000397.3)     | MAOB (monoamine oxidase B; NM_000898.4)                              | GOS2 (G0/G1 switch 2; NM_015714.3)                                             |
|  |  | IL5RA (interleukin 5 receptor subunit alpha; NM_001243099.1)                              | DNASE2B (deoxyribonuclease 2 beta; NM_058248.1)     | MLPH (melanophilin; NM_001042467.2)                                  | HIST1H2BC (histone cluster 1 H2B family member c; NM_003526.2)                 |
|  |  | KBTD11 (kelch repeat and BTB domain containing 11; NM_014867.2)                           | EMP1 (epithelial membrane protein 1; NM_001423.2)   | MPO (myeloperoxidase; NM_000250.1)                                   | HPSE (heparanase; NM_001098540.2)                                              |
|  |  | KCNH2 (potassium voltage-gated channel, subfamily H (eag-related), member 2; NM_000238.3) | FDX1 (ferredoxin 1; NM_004109.4)                    | MS4A2 (membrane spanning 4-domains A2; NM_001256916.1)               | IL8RA (interleukin 8 receptor alpha; L19591.1)                                 |
|  |  | LRP5L (LDL receptor related protein 5 like; NM_001135772.1)                               | FN1 (fibronectin 1; NM_001306131.1)                 | NM_003293 (tryptase alpha/beta 1; NM_003294.3)                       | IL8RB (interleukin-8 receptor type B; U11878.1)                                |
|  |  | MYO15B (myosin XVB; NM_001309242.1)                                                       | GM2A (GM2 ganglioside activator; NM_000405.4)       | NROB1 (nuclear receptor subfamily 0 group B member 1; NM_000475.4)   | KCNJ15 (potassium voltage-gated channel subfamily J member 15; NM_001276438.1) |

|  |  |                                                             |                                                                     |                                                                           |                                                                     |
|--|--|-------------------------------------------------------------|---------------------------------------------------------------------|---------------------------------------------------------------------------|---------------------------------------------------------------------|
|  |  | RCOR3 (REST corepressor 3; NM_001136224.2)                  | GPC4 (glypican 4; NM_001448.2)                                      | PGDS (hematopoietic prostaglandin D synthase; NM_014485.2)                | KIAA0329 (tectonin beta-propeller repeat containing 2; NM_014844.4) |
|  |  | RNASE2 (ribonuclease A family member 2; NM_002934.2)        | KAL1 (anosmin 1; NM_000216.3)                                       | PPM1H (protein phosphatase, Mg2+/Mn2+ dependent 1H; NM_020700.1)          | LILRB2 (leukocyte immunoglobulin like receptor B2; NR_103521.2)     |
|  |  | RNU2 (U2 snRNA; U57614.1)                                   | MARCO (macrophage receptor with collagenous structure; NM_006770.3) | PRG2 (proteoglycan 2, pro eosinophil major basic protein; NM_001302927.1) | MGAM (maltase-glucoamylase; NM_004668.2)                            |
|  |  | RRP12 (ribosomal RNA processing 12 homolog; NM_001284337.1) | ME1 (malic enzyme 1; NM_002395.5)                                   | PTGS1 (prostaglandin-endoperoxide synthase 1; NM_000962.3)                | MME (membrane metalloendopeptidase; NM_007289.2)                    |
|  |  | SIAH1 (siah E3 ubiquitin protein ligase 1; NM_003031.3)     | MS4A4A (membrane spanning 4-domains A4A; NM_001243266.1)            | SCG2 (secretogranin II; NM_003469.4)                                      | PDE4B (phosphodiesterase 4B; NM_001297440.1)                        |
|  |  | SMPD3 (sphingomyelin phosphodiesterase 3; NM_018667.3)      | MSR1 (macrophage scavenger receptor 1; NM_138716.2)                 | SIGLEC6 (sialic acid binding lg like lectin 6; NM_198845.5)               | S100A12 (S100 calcium binding protein A12; NM_005621.1)             |
|  |  | SYNJ1 (synaptojanin 1; NM_001160302.1)                      | PCOLCE2 (procollagen C-endopeptidase enhancer 2; NM_013363.3)       | SLC18A2 (solute carrier family 18 member A2; NM_003054.4)                 | SIGLEC5 (sialic acid binding lg like lectin 5; NM_003830.3)         |
|  |  | TGIF1 (TGFB induced factor homeobox 1; NM_001160302.1)      | PTGDS (prostaglandin D2 synthase; NM_001160302.1)                   | SLC24A3 (solute carrier family 24 member A3; NM_001160302.1)              | SLC22A4 (solute carrier family 22 member 4; NM_001160302.1)         |

|  |  |                                                                               |                                                                       |                                                            |                                                                        |
|--|--|-------------------------------------------------------------------------------|-----------------------------------------------------------------------|------------------------------------------------------------|------------------------------------------------------------------------|
|  |  | NM_174886.2)                                                                  | NM_000954.5)                                                          | member 3;<br>NM_020689.3)                                  | NM_003059.2)                                                           |
|  |  | THBS1<br>(thrombospondin 1;<br>NM_003246.3)                                   | RAI14 (retinoic acid<br>induced 14;<br>NM_001145525.1)                | TAL1 (T-cell acute<br>lymphocytic leukemia<br>1; X51990.1) | SLC25A37 (solute<br>carrier family 25<br>member 37;<br>NM_001317812.1) |
|  |  | THBS4<br>(thrombospondin 4;<br>NM_001306213.1)                                | SCARB2 (scavenger<br>receptor class B<br>member 2;<br>NM_001204255.1) | TPSAB1 (tryptase<br>alpha/beta 1;<br>NM_003294.3)          | TNFRSF10C (TNF<br>receptor superfamily<br>member 10c;<br>NM_003841.3)  |
|  |  | TIPARP (TCDD<br>inducible poly(ADP-<br>ribose) polymerase;<br>NM_001184718.1) | SCG5 (secretogranin<br>V; NM_001144757.2)                             | TPSB2 (tryptase beta<br>2; NM_024164.5)                    | VNN3 (vanin 3;<br>NM_001291703.1)                                      |
|  |  | TKTL1 (transketolase<br>like 1;<br>NM_001145934.1)                            | SGMS1<br>(sphingomyelin<br>synthase 1;<br>NM_147156.3)                |                                                            |                                                                        |
|  |  |                                                                               | SULT1C2<br>(sulfotransferase<br>family 1C member 2;<br>NM_176825.2)   |                                                            |                                                                        |

\*Each GenBank Accession Number is a representative or exemplary GenBank Accession Number for the listed gene and is herein incorporated by reference in its entirety for all purposes. Further, each listed representative or exemplary accession number should not be construed to limit the claims to the specific accession number.

**[00175] Table 5.** Individual Immunomarkers for use in the methods provided herein.

| Gene Name                                   | Abbreviation | GenBank Accession No.* |
|---------------------------------------------|--------------|------------------------|
| Programmed Death Ligand 1                   | PDL1         | NM_014143              |
| programmed death ligand 2                   | PDL2         | AY254343               |
| programmed cell death 1                     | PDCD1        | NM_005018              |
| cytotoxic T-lymphocyte associated protein 4 | CTLA4        | NM_005214              |

\*Each GenBank Accession Number is a representative or exemplary GenBank Accession Number for the listed gene and is herein incorporated by reference in its entirety for all purposes. Further, each listed representative or exemplary accession number should not be construed to limit the claims to the specific accession number.

**[00176] Table 6.** Interferon (IFN) Genes for use in the methods provided herein.

| Gene Name                                                   | Abbreviation | GenBank Accession No.* |
|-------------------------------------------------------------|--------------|------------------------|
| Chemokine (C-X-C Motif) Ligand 10                           | CXCL10       | NM_001565              |
| C-X-C motif chemokine ligand 9                              | CXCL9        | NM_002416              |
| interferon alpha inducible protein 27                       | IFI27        | NM_001130080           |
| interferon induced protein with tetratricopeptide repeats 1 | IFIT1        | NM_001548              |
| interferon induced protein with tetratricopeptide repeats 2 | IFIT2        | NM_001547              |
| interferon induced protein with tetratricopeptide repeats 3 | IFIT3        | NM_001549              |
| MX dynamin like GTPase 1                                    | MX1          | NM_001144925           |
| MX dynamin like GTPase 2                                    | MX2          | XM_005260983           |
| 2'-5'-oligoadenylate synthetase 1                           | OAS1         | NM_016816              |
| 2'-5'-oligoadenylate synthetase 2                           | OAS2         | NM_016817              |
| signal transducer and activator of transcription            | STAT1        | NM_007315              |

| Gene Name                                        | Abbreviation | GenBank Accession No.* |
|--------------------------------------------------|--------------|------------------------|
| 1                                                |              |                        |
| signal transducer and activator of transcription | STAT2        | NM_005419              |
| 2                                                |              |                        |

\*Each GenBank Accession Number is a representative or exemplary GenBank Accession Number for the listed gene and is herein incorporated by reference in its entirety for all purposes. Further, each listed representative or exemplary accession number should not be construed to limit the claims to the specific accession number.

[00177] **Table 7.** MHC class II genes for use in the methods provided herein.

| Name     | Abbreviation                                                        | GenBank Accession No.* |
|----------|---------------------------------------------------------------------|------------------------|
| CD74     | Homo sapiens CD74 molecule (CD74)                                   | NM_001025159           |
| CIITA    | class II major histocompatibility complex transactivator            | NM_001286402           |
| CTSH     | cathepsin H                                                         | NM_004390              |
| HLA-DMA  | Homo sapiens major histocompatibility complex, class II, DM alpha   | NM_006120              |
| HLA-DPA1 | Homo sapiens major histocompatibility complex, class II, DP alpha 1 | NM_033554              |
| HLA-DPB1 | Human MHC class II lymphocyte antigen (HLA-DP) beta chain           | M83664                 |
| HLA-DQA1 | Homo sapiens major histocompatibility complex, class II, DQ alpha 1 | NM_002122              |
| HLA-DRB1 | Homo sapiens major histocompatibility complex, class II, DR beta 1  | NM_002124              |
| HLA-DRB5 | Homo sapiens major histocompatibility complex, class II, DR beta 5  | NM_002125              |
| HLA-DRB6 | Homo sapiens major histocompatibility complex, class II, DR beta 6  | NR_001298              |
| NCOA1    | Homo sapiens nuclear receptor coactivator 1                         | NM_003743              |

\*Each GenBank Accession Number is a representative or exemplary GenBank Accession Number for the listed gene and is herein incorporated by reference in its entirety for all purposes. Further, each listed representative or exemplary accession number should not be construed to limit the claims to the specific accession number.

**[00178]** In one embodiment, upon determining a patient's SQ lung cancer subtype using any of the methods and classifier biomarkers panels or subsets thereof as provided herein alone or in combination with determining expression of one or more immune cell markers as provided herein, the patient is selected for treatment with or administered an immunotherapeutic agent. The immunotherapeutic agent can be a checkpoint inhibitor, monoclonal antibody, biological response modifiers, therapeutic vaccine or cellular immunotherapy.

**[00179]** In another embodiment, the immunotherapeutic agent is a checkpoint inhibitor. In some cases, a method for determining the likelihood of response to one or more checkpoint inhibitors is provided. In one embodiment, the checkpoint inhibitor is a PD-1/PD-L1 checkpoint inhibitor. The PD-1/PD-L1 checkpoint inhibitor can be nivolumab, pembrolizumab, atezolizumab, durvalumab, lambrolizumab, or avelumab. In one embodiment, the checkpoint inhibitor is a CTLA-4 checkpoint inhibitor. The CTLA-4 checkpoint inhibitor can be ipilimumab or tremelimumab. In one embodiment, the checkpoint inhibitor is a combination of checkpoint inhibitors such as, for example, a combination of one or more PD-1/PD-L1 checkpoint inhibitors used in combination with one or more CTLA-4 checkpoint inhibitors.

**[00180]** In one embodiment, the immunotherapeutic agent is a monoclonal antibody. In some cases, a method for determining the likelihood of response to one or more monoclonal antibodies is provided. The monoclonal antibody can be directed against tumor cells or directed against tumor products. The monoclonal antibody can be panitumumab, matuzumab, necitumumab, trastuzumab, amatuximab, bevacizumab, ramucirumab, bavituximab, patritumab, rilotumumab, cetuximab, immu-132, or demcizumab.

**[00181]** In yet another embodiment, the immunotherapeutic agent is a therapeutic vaccine. In some cases, a method for determining the likelihood of response to one or more therapeutic vaccines is provided. The therapeutic vaccine can be a peptide or tumor cell vaccine. The vaccine can target MAGE-3 antigens, NY-ESO-1 antigens, p53 antigens, survivin antigens, or MUC1 antigens. The therapeutic cancer vaccine can be GVAX (GM-CSF gene-transfected tumor cell vaccine), belagenpumatucel-L (allogeneic tumor cell vaccine made with four irradiated NSCLC cell lines modified with TGF-beta2 antisense plasmid), MAGE-A3 vaccine (composed of MAGE-A3 protein and adjuvant AS15), (I)-BLP-25 anti-MUC-1 (targets MUC-1 expressed on tumor cells), CimaVax EGF (vaccine

composed of human recombinant Epidermal Growth Factor (EGF) conjugated to a carrier protein), WT1 peptide vaccine (composed of four Wilms' tumor suppressor gene analogue peptides), CRS-207 (live-attenuated *Listeria monocytogenes* vector encoding human mesothelin), Bec2/BCG (induces anti-GD3 antibodies), GV1001 (targets the human telomerase reverse transcriptase), tergenpumatucl-L (consists of human lung cancer cells genetically modified to include a mouse gene to which the immune system responds strongly), TG4010 (targets the MUC1 antigen), racotumomab (anti-idiotypic antibody which mimics the NGcGM3 ganglioside that is expressed on multiple human cancers), tecemotide (liposomal BLP25; liposome-based vaccine made from tandem repeat region of MUC1) or DRibbles (a vaccine made from nine cancer antigens plus TLR adjuvants).

**[00182]** In one embodiment, the immunotherapeutic agent is a biological response modifier. In some cases, a method for determining the likelihood of response to one or more biological response modifiers is provided. The biological response modifier can trigger inflammation such as, for example, PF-3512676 (CpG 7909) (a toll-like receptor 9 agonist), CpG-ODN 2006 (downregulates Tregs), *Bacillus Calmette-Guerin* (BCG), *Mycobacterium vaccae* (SRL172) (nonspecific immune stimulants now often tested as adjuvants). The biological response modifier can be cytokine therapy such as, for example, IL-2+ tumor necrosis factor alpha (TNF-alpha) or interferon alpha (induces T-cell proliferation), interferon gamma (induces tumor cell apoptosis), or Mda-7 (IL-24) (Mda-7/IL-24 induces tumor cell apoptosis and inhibits tumor angiogenesis). The biological response modifier can be a colony-stimulating factor such as, for example granulocyte colony-stimulating factor. The biological response modifier can be a multi-modal effector such as, for example, multi-target VEGFR: thalidomide and analogues such as lenalidomide and pomalidomide, cyclophosphamide, cyclosporine, denileukin diftitox, talactoferrin, trabectedin or all-trans-retinoic acid.

**[00183]** In one embodiment, the immunotherapy is cellular immunotherapy. In some cases, a method for determining the likelihood of response to one or more cellular therapeutic agents. The cellular immunotherapeutic agent can be dendritic cells (DCs) (ex vivo generated DC-vaccines loaded with tumor antigens), T-cells (ex vivo generated lymphokine-activated killer cells; cytokine-induced killer cells; activated T-cells; gamma delta T-cells), or natural killer cells.

**[00184]** In some cases, specific subtypes of SQ have different levels of immune activation (e.g., innate immunity and/or adaptive immunity) such that subtypes with elevated or detectable immune activation (e.g., innate immunity and/or adaptive immunity) are selected for treatment with one or more immunotherapeutic agents described herein. In one embodiment, the classical subtype of SQ has low immune activation as compared to other SQ subtypes or lung cancer subtypes. In some cases, specific subtypes of SQ have high or elevated levels of immune activation. In some cases, the secretory subtype of SQ has elevated levels of immune activation (e.g., innate immunity and/or adaptive immunity) as compared to other SQ subtypes or lung cancer subtypes. In one embodiment, SQ subtypes with low levels of or no immune activation (e.g., innate immunity and/or adaptive immunity) are not selected for treatment with one or more immunotherapeutic agents described herein.

### **Detection Methods**

**[00185]** In one embodiment, the methods and compositions provided herein allow for the detection of at least one nucleic acid in a lung cancer sample (e.g. squamous cell carcinoma lung cancer sample) obtained from a subject. The at least one nucleic acid can be a classifier biomarker provided herein. In one embodiment, the at least one nucleic acid detected using the methods and compositions provided herein are selected from **Table 1**. In one embodiment, the methods of detecting the nucleic acid(s) (e.g., classifier biomarkers) in the lung cancer sample obtained from the subject comprises, consists essentially of, or consists of measuring the expression level of at least one or a plurality of biomarkers using any of the methods provided herein. The biomarkers can be selected from **Table 1**. In some cases, the plurality of biomarker nucleic acids comprises, consists essentially of or consists of at least two biomarker nucleic acids, at least 10 biomarker nucleic acids, at least 20 biomarker nucleic acids, at least 30 biomarker nucleic acids, at least 40 biomarker nucleic acids, at least 50 biomarker nucleic acids, at least 60 biomarker nucleic acids, at least 70 biomarker nucleic acids or all 80 biomarkers nucleic acids of **Table 1**. The detection can be at the nucleic acid level. The detection can be by using any amplification, hybridization and/or sequencing assay disclosed herein.

**[00186]** In another embodiment, the methods and compositions provided herein allow for the detection of at least one nucleic acid or a plurality of nucleic acids in a lung cancer sample (e.g. squamous cell carcinoma lung cancer sample) obtained from a subject such that the at least one nucleic acid is or the plurality of nucleic acids are selected from the biomarkers listed in **Table 1** and the detection of at least one biomarker from a set of biomarkers whose presence, absence and/or level of expression is indicative of immune activation. The set of biomarkers for indicating immune activation can be gene expression signatures of and/or Adaptive Immune Cells (AIC) (e.g., **Table 4A**) and/or Innate Immune Cells (IIC) (e.g., **Table 4B**), individual immune biomarkers (e.g., **Table 5**), interferon genes (e.g., **Table 6**), major histocompatibility complex, class II (MHC II) genes (e.g., **Table 7**) or a combination thereof. The gene expression signatures of both IIC and AIC can be any gene signatures known in the art such as, for example, the gene signature listed in Bindea et al. (Immunity 2013; 39(4); 782-795). The detection can be at the nucleic acid level. The detection can be by using any amplification, hybridization and/or sequencing assay disclosed herein.

### **Kits**

**[00187]** Kits for practicing the methods of the invention can be further provided. By "kit" is intended any manufacture (e.g., a package or a container) comprising at least one reagent, e.g., an antibody, a nucleic acid probe or primer, etc., for specifically detecting the expression of a biomarker of the invention. The kit may be promoted, distributed, or sold as a unit for performing the methods of the present invention. Additionally, the kits may contain a package insert describing the kit and methods for its use.

**[00188]** In one embodiment, kits for practicing the methods of the invention are provided. Such kits are compatible with both manual and automated immunocytochemistry techniques (e.g., cell staining). These kits comprise at least one antibody directed to a biomarker of interest, chemicals for the detection of antibody binding to the biomarker, a counterstain, and, optionally, a bluing agent to facilitate identification of positive staining cells. Any chemicals that detect antigen- antibody binding may be used in the practice of the invention. The kits may comprise at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, or more antibodies for use in the methods of the invention.

## **EXAMPLES**

[00189] The present invention is further illustrated by reference to the following Examples. However, it should be noted that these Examples, like the embodiments described above, is illustrative and is not to be construed as restricting the scope of the invention in any way.

### **Example 1 – Immune Cell Activation Differences Among Lung Squamous Cell Carcinoma Intrinsic Subtypes and Variable Correlation With CD274 (PD-L1) Expression.**

#### **Introduction**

[00190] Gene expression based subtyping in Lung Squamous Cell Carcinoma (SQ) classifies SQ tumors into distinct subtypes with variable biologic and clinical features. Gene expression based subtyping has consistently identified 4 subtypes with Lung SQ, Primitive, Classical, Basal and Secretory (1, 2) (see **FIG. 1**). SQ subtypes demonstrate key differences in genomic alterations, tumor drivers, prognosis, and likely response to various therapies (1-2).

#### **Methods**

[00191] Using previously published Bindea et al. (3) immune cell gene signatures (24 in total) and SQ subtyping gene expression signatures (1-2), several publically available lung SQ datasets (1-2 and 4-5) (see **FIG. 2**) were examined for immune cell features in relation to SQ subtypes. This investigation of immune differences by subtype used the 24 immune cell gene signatures from Bindea et al [3] that each had a varying number of genes and were classified as adaptive or innate immunity cell signatures (*see Table 4A-4B*). Adaptive Immune Cell (AIC) signatures (**Table 4A**) included Tcells, Central Memory T cells (Tcm), Effector Memory T cells (Tem), T helper cell (Th), Type 1 T helper cells (Th1), Type 2 T helper cells (Th2), T follicular helper cells (Tfh), T helper 17 cells (Th17), T Regulatory Cells (Treg), Gamma Delta T cells (Tgd), CD8 Tcells, Cytotoxic T cells, B cells, and Innate Immune Cell (IIC) signatures (**Table 4B**) included Natural Killer (NK), NK CD56dim cells, NK CD56bright cells, Dendritic cells (DC), Immature Dendritic Cells (iDC), Dendritic Cells (pDC), Activated Dendritic Cells(aDC), Mast cells, Eosinophils, Macrophages, and Neutrophils. In addition to the gene expression signatures of both Innate Immune Cells (IIC) and

Adaptive Immune Cells (AIC), a 13 gene IFN signature (IFN; **Table 6**), a 13-gene MHC class II signature score (Forero [6]; **Table 7**) as well as single gene immune biomarkers in **Table 5** (CTLA4, PDCD1, CD274 (PD-L1), and PDCDLG2 (PD-L2)) were examined in the 4 SQ subtypes (Primitive, Classical, Secretory, Basal).

**[00192]** For SQ, 4 published gene expression data sets of lung squamous cell carcinoma samples having a total of 762 patients were used, including TCGA [2], Lee et al [4], Raponi et al [5], and Wilkerson et al [1]. For TCGA, upper quantile normalized RSEM data was downloaded from Firehose and log2 transformed. Normalized Affy array data was downloaded from GEO for Lee et al [4] (GSE8894) and Raponi et al [5] (GSE4573), and normalized Agilent array data was downloaded from GEO for Wilkerson et al [1] (GSE17710).

**[00193]** To determine the squamous cell carcinoma subtype (basal, classical, primitive, secretory), the published 208-gene nearest centroid classifier as described previously in Wilkerson et al [1] was used. After median centering of genes in the signature, each sample was assigned the subtype corresponding to the centroid with which it was maximally correlated. (Pearson)

**[00194]** Using the TCGA data for squamous cell carcinoma, the correlations were assessed among the 30 markers by plotting matrices of pairwise Spearman rank correlation coefficients where markers were ordered by hierarchical clustering (see **FIG. 4**). To investigate overall immunity marker trends by subtype, the expression heatmaps were plotted where samples were arranged by subtype and markers were grouped according to ordering in Bindea et al [3] (see **FIG. 3**). To evaluate the reproducibility of immunity marker differences among the subtypes, normalized T cell signatures were plotted by subtype for each data set (see **FIG. 5**).

**[00195]** Immune cell signature associations with tumor subtype and with CD274 expression were evaluated using linear regression. More specifically, to assess the prediction strength of subtype as a predictor of immune markers relative to that of PD-L1, a linear regression model of each signature with subtype the sole predictor, and again with PD-L1 the sole predictor, was fitted in the TCGA dataset. PD-L1 expression was treated as a low/medium/high categorical variable with equal proportions in each group. Scatter plots of adjusted R-squared when subtype was the predictor against adjusted R-squared when PD-L1 was the predictor were inspected for overall trends (see **FIG. 6**).

[00196] Using non-silent mutation burden per Mb data, available in the supplementary information from TCGA squamous cell carcinoma (Lawrence 2013), mutation burden-Tcell expression associations was investigated using the Kruskal Wallis test and the Spearman correlation coefficients, respectively. For TCGA squamous cell carcinoma, NFE2L2-subtype association was evaluated using the Kruskal Wallis test. To test whether NFE2L2 in SQ showed evidence of association after adjusting for subtype, a linear model for Tcell expression was fit with NFE2L2 expression in SQ as sole predictors and again following adjustment for subtype.

[00197] Subtype and immune signature associations with a 13-gene MHC class II signature [Forero [6]; **Table 7**, calculated as an average of all genes in the list (**Table 7**), were investigated using the Kruskal-Wallis test. For immune signature-MHC class II associations, Spearman correlation coefficients were calculated.

[00198] Hierarchical clustering of immune signatures and pairwise signature correlations were also analyzed. Survival signature associations of Stages I-III samples were evaluated with stratified cox proportional hazard models allowing for different baseline hazards in each dataset. More specifically, immune marker-survival associations in the TCGA data sets were tested, overall and separately within each subtype, using Cox proportional hazards models. Immune markers were centered and scaled to have mean 0 and variance 1, and stage IV patients were excluded. Evaluations within a specific subtype adjusted for stage, and overall evaluations adjusted for both stage and subtype. Forest plots showing hazard ratios and confidence intervals for each signature were made (see **FIGs. 7A** and **7B**). All statistical analyses were conducted using R 3.2.0 software (<http://www.R-project.org>).

## Results

[00199] Heatmap analysis and unsupervised hierarchical clustering of immune cell gene signatures provided separation of intrinsic subtypes of SQ (see **FIGs. 3** and **4**). Examination of Immune cell gene signatures (both AIC and IIC) as well as individual immune gene markers revealed clear differences among the SQ subtypes (see **FIG. 3**).

[00200] Among the SQ subtypes, the secretory subtype showed consistently higher immune cell expression of both innate and adaptive immune cells with one exception, the Th2 signature, where both primitive and secretory had comparable expression (**FIG. 21**). The classical subtype demonstrated the lowest immune cell expression of all the SQ subtypes. Unlike the case for AD subtypes, *CD274* (PD-L1) expression did not correlate with other immune cell expression in SQ

subtypes. This is especially obvious in the classical subtype where CD274 (PD-L1) expression was high despite relatively low expression of other immune cells (see **FIG. 3** and **FIG. 21**). Overall, immune activation was most prominent in the secretory subtype of SQ demonstrating activation of both innate as well as adaptive immune cells. In contrast, the classical subtype of SQ demonstrated lower immune activation.

[00201] Using hierarchical clustering, correlation matrices revealed clustering of adaptive immune cells and innate immune cells (see **FIG. 4**). In SQ, adaptive immune features such as T cells, cytotoxic cells, CD8 cells, Th1 cells, *PDCD1*, *CTLA4*, and Tregs had high pairwise correlations and similarly for innate immune cells, including iDC, DC, macrophages, neutrophils, mast cells, and eosinophils are correlated (**FIG. 4**). Further, in SQ, NK CD56dim cells (cytolytic activity) were more strongly correlated with adaptive immune cells than with innate immune cells (see **FIG. 4**). In addition, TFH and B cells were more highly correlated with adaptive immune features in SQ (see **FIG. 4**).

[00202] Strength of association of *CD274* (PD-L1) expression with adaptive immune cell signatures, as compared to SQ subtype was conducted. As shown in **FIG. 6**, in SQ tumors, subtype was a better predictor of immune cell expression than *CD274* (PD-L1) expression for all adaptive immune cells examined (median F-test p-value and adjusted R-squared were  $2.16\text{e-}24$  and 0.20 for subtype versus  $1.86\text{e-}09$  and 0.07 for *CD274*).

[00203] Immune cell signatures were primarily evaluated in the TCGA datasets, however SQ subtype immune differences, as measured by the immune cell signatures, were found to be very reproducible across multiple datasets (see **FIG. 5**). T cell immune cell signature expression subtype differences in SQ subtypes were remarkably reproducible across a variety of gene expression datasets derived from both frozen and FFPE samples and involving a variety of gene expression platforms including RNAseq (Illumina, San Diego, CA) and microarrays from both Affymetrix (Santa Clara, CA) and Agilent (Santa Clara, CA). Overall, immune cell signature gene expression patterns were consistent across multiple SQ (see **FIG. 5**) datasets.

[00204] In SQ, non-silent mutation burden was not significantly different across subtypes (see **FIG. 22**). Mutation burden was not strongly correlated with Tcell immune cell expression in SQ datasets (Spearman correlation = -0.08 in SQ).

[00205] Several other genomic features such as *KEAP1/NFE2L2* alterations in SQ (Hast [7]) have been suggested as possible contributors to reduced immune response in NSCLC.

*KEAP/NFE2L2* alterations, impacting the oxidative stress pathway, were enriched in the SQ classical subtype (**FIG. 22**). Alterations in the cyclooxygenase 2 (*COX2*) pathway as measured by increased *NFE2L2* expression in SQ were associated with lower immune cell expression, however after adjustment for subtype using linear regression, *NFE2L2* were significant predictors (*NFE2L2* expression in SQ  $p=1.2E-07$  to  $p=0.47$  following adjustment for subtype).

**[00206]** The association of immune cell expression in SQ lung cancer with MHC class II genes was investigated using a published 13 gene MHC class II signature (Forero [6]). MHC class II gene expression was strongly correlated with several immune cells in SQ including Tcell expression (Spearman correlation=0.86 in SQ), Bcell expression (Spearman correlation=0.69 in SQ) and DC expression (Spearman correlation=0.76 in SQ). MHC class II gene expression was significantly higher in tumor adjacent normal lung tissue as compared with tumor and was differentially expressed across tumor subtypes (**FIG. 22**). In a linear model of the MHC class II signature as a predictor of Tcell immune cell expression, MHC class II remained significant following adjustment for SQ subtype ( $p<1E-50$  for MHC II).

**[00207]** Using cox proportional hazard models, subtype specific hazard ratios (HRs) for one unit of increased expression were calculated. Subtype specific HRs were adjusted for pathologic stage and confidence intervals (CI) were calculated. Hazard ratios and confidence intervals for markers that were significant (nominal  $p\text{-value}<0.05$ ) for at least one subtype are shown in **FIGs. 7A-7B**. The HR and CI for cell signatures or genes showing significant survival associations for one or more of the subtypes are shown in **FIGs. 7A-7B**. Among the SQ subtypes, a unit increase in expression of Th1, Th2, TFH, DC, macrophages, mast cells, and MHC class II was significantly associated with improved survival in the primitive subtype (**FIGs. 7A-7B**). Curiously, the secretory subtype did not show significant association with survival possibly due to the uniformly high expression of immune cells in the secretory subtype preventing demonstration of an incremental survival benefit per unit increase. In SQ, only the primitive subtype demonstrated significant immune cell expression associations with improved survival ( $p<0.01$ ) (see **FIG. 7A-7B**).

## Conclusion

**[00208]** Lung SQ gene expression subtypes vary in their immune landscape. Intrinsic biologic subtypes of SQ reveal key differences in immune cell activation, which were not always correlated with *CD274* expression and demonstrated variable association with survival. SQ

Classical subtype showed minimal immune infiltration (depressed immune cell expression) suggesting reduced response to immunoRX, while the secretory subtype showed elevated immune expression among the SQ tumor subtypes. In SQ, subtype appeared to be a better predictor of immune infiltration than CD274 (PD-L1). CD274 expression was not associated with AIC expression nor with improved survival in SQ. The SQ primitive subtype showed immune feature expression associated with improved survival. Further, non-silent mutation burden was not correlated with immune cell expression across subtypes; however, MHC class II gene expression was highly correlated. Increased immune and MHC II gene expression was associated with improved survival in the primitive subtype of SQ.

### **Incorporation by reference**

[00209] The following references are incorporated by reference in their entireties for all purposes.

[00210] 1.) Wilkerson MD, et al. Clin Cancer Res 2010; 16(19): 4864-75. PMID 20643781

[00211] 2.) TCGA Lung SQCC. Nature 2012; 489(7417): 519-525. PMID 22960745

[00212] 3.) Bindea et al., Immunity 2013; 39(4): 782-95. PMID 24138885

[00213] 4.) Lee ES, et al. Cancer Res 2008; 14(22): 7397-7404. PMID 19010856

[00214] 5.) Raponi M, et al. Cancer Res 2006; 66(7): 466-72. PMID 16885343

[00215] 6.) Forero A, Li Y, Dongquan C, et al. Expression of the MHC class II pathway in triple negative breast cancer tumor cells is associated with a good prognosis and infiltrating lymphocytes. Cancer Immunol Res 2016; 4(5):390-399.

[00216] 7.) Hast BE, Cloer EW, Goldfarb D, et al. Cancer-derived mutations in KEAP1 impair NRF2 degradation but not ubiquitination. Cancer Res 2014; 74(3): 808-817.

## **EXAMPLE 2 Development and Validation of the Lung Squamous Cell Carcinoma**

### **Subtyping Signature**

#### **Objective**

[00217] Lung squamous cell carcinoma (SQ) subtyping has been primarily restricted to a research protocol involving the extraction of RNA from Fresh Frozen lung tumors, followed by application of a nearest centroid predictor using quantitative gene expression of over 200 genes.

Despite evidence of prognostic and predictive benefits from squamous cell carcinoma subtyping, the need for Fresh Frozen tissue, the requirement for gene expression of >200 genes in combination with complex bioinformatic analyses, has hindered the application of SQ subtyping in drug development and/or the clinic. The goal of this study was to develop a robust and efficient gene signature (with fewer genes needed) for differentiating the four subtypes of squamous cell carcinoma (*i.e.*, basal, classical, secretory or primitive subtypes). The new efficient gene signature may serve to reliably subtype SQ from fresh frozen or FFPE tumor samples, making it amenable for diagnostic applications and/or drug development using any of the available quantitative RNA platforms (qRT-PCR, RNAseq, Affymetrix or Agilent Arrays). Development of the 80 gene signature for differentiating the subtypes of squamous cell carcinoma was performed as described in the methods herein.

## Methods

**[00218]** Employing a Classifying arrays to Nearest Centroid (CLaNC) [1] algorithm, a number of approaches were tested to determine an optimal number of genes to include in a SQ subtyping gene signature. In one approach (plain approach in **FIG. 8**), the CLaNC was applied to the TCGA lung SQ RNAseq gene expression dataset (n=501) without modification. In a second approach (50-50 high low in **FIGs. 8** and **9**), the CLaNC was used on the TCGA lung SQ RNAseq gene expression dataset (n=501) with modification to select an equal number of negatively and positively correlated genes for each SQ subtype. In a third approach (50-50 but 20k in **FIG. 8**), the CLaNC was used on the transcriptome dataset (n=20,000) with modification to select an equal number of negatively and positively correlated genes for each SQ subtype. In a final approach (hybrid approach in **FIG. 8**), the CLaNC was used on the TCGA lung SQ RNAseq gene expression dataset (n=501) with modification to select an equal number of negatively and positively correlated genes for the basal and secretory SQ subtypes, a number of negatively correlated gene for the primitive SQ subtype and a number of positively correlated genes for the classical SQ subtype. Examination of the optimal number of genes to include in the SQ signature was chosen based on evaluation of 5-fold cross validation curves for each approach described above (*see* **FIG. 8**). Ultimately, the 50-50 high low approach was selected as the approach for determining an optimal number of genes to include in the SQ subtyping gene

signature. This approach showed that examination of the expression patterns of 20 genes per subtype or 80 total genes could be used to accurately subtype a SQ sample.

**[00219]** Using the TCGA lung SQ RNAseq gene expression dataset (n=501) for training and the 208-gene classifier to define gold standard subtype, an 80-gene signature was developed that maintains low misclassification rates when applied to several independent test sets. Starting with the standard 208 classifier genes, the Classifying arrays to Nearest Centroid (CLaNC) [1] algorithm was used with modification to select an equal number of negatively and positively correlated genes for each subtype as described above. The optimal number of genes (20 per subtype) to include in the signature was chosen based on 5-fold cross validation curves performed using the TCGA lung SQ dataset (see **FIGs. 8** and **9**). Selection of prototype samples (**FIG. 10**) for training of the predictor entailed applying the CLaNC to the entire TCGA data set minus 20% of samples with the lowest gold standard subtype prediction strength, removing an equal number from each subtype. The gene rank of genes from the 80-gene signature in the gold standard 208 gene classifier [2] can be seen in **FIG. 11**.

**[00220]** The 80-gene signature was then tested in several Fresh Frozen publicly available array and RNAseq datasets [2, 3, 4, 5] and results were compared with the gold standard subtype calls as defined by the previously published 208-gene signature [2]. Final validation of the 80-gene signature (**Table 1**) was then performed in a newly collected RNAseq dataset of archived FFPE squamous cell carcinoma samples to assure comparable performance in FFPE samples.

**[00221]** In order to validate the consistent performance of the selected 80 gene signature, the newly collected FFPE samples were lung squamous cell carcinoma (SQ) residual archived samples (primarily surgical samples) that had been collected under an IRB approved protocol at the University of North Carolina in Chapel Hill, NC. The samples were reviewed by a pathologist for tumor cells and three 10µm tissue sections were macrodissected prior to extraction to enrich for tumor cells. RNA was quantitated and 100ng was input per sample. Sequencing libraries were constructed using Illumina RNA-Access kits that enrich for the transcriptome. Sequencing libraries were under quality control by using a BA analyzer and quantified using qPCR. Sequence data was generated on an Illumina HiSeq platform (50 bp PE, 20-30 million reads) and was under quality control by using fastQC. Sequence results were

aligned against hg19 reference sequence using STAR aligner and the transcriptome was built using Cufflinks [6]. Cuffcompare was used to annotate the transcriptome and counts of various expressed genes were calculated. RSEM expression count estimates were upper quartile normalized and log2 transformed following the approach used in the Cancer Genome Atlas lung squamous cell carcinoma analysis [3, 7].

## Results

**[00222]** The 80 gene signature gene list developed in this study is shown in **Table 2**, while the T statistics for the 80 gene signature gene list for each SQ subtype can be found in **Table 1**. The median gene expression of the 20 genes selected for each SQ subtype (basal, classical, primitive, or secretory) is shown in **FIGs. 12, 13, 14 and 15**, respectively. Agreement of subtype calls using the 80 gene signature with the published 208 gene signature subtype call in several different test datasets is shown in **FIG. 16**. The newly developed 80 gene signature demonstrated agreement in a range of 0.84-0.91 in the other 4 test datasets (**FIG. 16**) and the new collected FFPE samples (**FIG. 16**). Below is a summary of the test datasets, the types of the RNA platforms, and the numbers of the squamous cell carcinoma samples used.

| Reference                                | RNA Platform      | Squamous Cell Carcinoma Samples |
|------------------------------------------|-------------------|---------------------------------|
| TCGA Squamous Cell Carcinoma             | RNAseq            | 501                             |
| Lee                                      | Affymetrix Arrays | 75                              |
| Raponi                                   | Affymetrix Arrays | 130                             |
| UNC                                      | RNAseq            | 56                              |
| Newly collected GeneCentric FFPE samples | RNAseq            | 46                              |

## Conclusion

**[00223]** Development and validation of an efficient 80 gene signature for SQ subtyping was described. The resulting 80 gene signature maintains low misclassification rates when applied to several independent test sets. Thus, the new signature reliably subtypes SQ from fresh frozen or FFPE tumor samples and can perform reliably using gene expression data generated from a variety of platforms including RNAseq and Arrays.

**Incorporation by reference**

[00224] The following references are incorporated by reference in their entireties for all purposes.

[00225] 1.) Dabney AR. ClaNC: Point-and-click software for classifying microarrays to nearest centroids. *Bioinformatics*. 2006;22: 122–123. doi:10.1093/bioinformatics/bti756

[00226] 2.) Wilkerson MD, et al. *Clin Cancer Res* 2010; 16(19): 4864-75. PMID 20643781

[00227] 3.) TCGA Lung SQCC. *Nature* 2012; 489(7417): 519-525. PMID 22960745

[00228] 4.) Lee ES, et al. *Cancer Res* 2008; 14(22): 7397-7404. PMID 19010856

[00229] 5.) Raponi M, et al. *Cancer Res* 2006; 66(7): 466-72. PMID 16885343

[00230] 6.) Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature biotechnology* 2010;28(5):511-5.

[00231] 7.) Li B, and Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 2011, 12:323 doi:10.1186/1471-2105-12-323

**Example 3: Immune Cell Activation Differences Among Lung Squamous Cell Carcinoma Intrinsic Subtypes as determined using Lung Squamous Cell Carcinoma Subtyping 80 Gene Signature from Example 2.**

**Methods**

[00232] Using previously published Bindea et al. (1) immune cell gene signatures (24 in total) and the Lung SQ subtyping gene signature described in Example 2 for subtyping SQ, several publically available lung SQ datasets (2-5; see FIG. 2), were examined for immune cell features in relation to SQ subtypes. Gene expression signatures of both Innate Immune Cells (IIC) and Adaptive Immune Cells (AIC), a 13 gene IFN signature (IFN), as well as single gene immune biomarkers (CTLA4, PDCD1, and CD274 (PD-L1), PDCDLG2 (PD-L2)) were examined in the 4 SQ subtypes (basal, classical, primitive and secretory). Immune cell signature associations with tumor subtype and with CD274 expression were evaluated using linear regression. Hierarchical clustering of immune signatures and pairwise signature correlations were also analyzed. Survival

signature associations of Stages I-III samples were evaluated with stratified cox proportional hazard models allowing for different baseline hazards in each dataset.

## Results

[00233] Using the TCGA SQ dataset and the 80 gene SQ subtyping signature of Example 2, heatmap analysis and unsupervised hierarchical clustering of immune cell gene signatures provided separation of intrinsic subtypes of SQ in a similar fashion as to what was observed in Example 1 (see FIG. 3 and FIG. 17). Further, immune cell signature gene expression patterns were consistent across multiple SQ (see FIG. 18) datasets similar to that observed in Example 1 (see FIG. 5). As in Example 1, strength of association of *CD274* (PD-L1) expression with adaptive immune cell signatures, as compared to SQ subtype was conducted. As shown in FIG. 19 (like in FIG. 6), in SQ tumors, subtype was a better predictor of immune cell expression than *CD274* (PD-L1) expression for all adaptive immune cells examined.

[00234] Using cox proportional hazard models, subtype specific hazard ratios for one unit of increased expression were calculated as described in Example 1. Subtype specific HR's were adjusted for pathologic stage and confidence intervals were calculated. The HR and CI for cell signatures or genes showing significant survival associations for one or more of the subtypes are shown in FIG. 20. Among the SQ subtypes, a unit increase in expression of Th1, Th2, TFH, DC, macrophages, and mast cells was significantly associated with improved survival in the primitive subtype much like in Example 1 (see FIGs. 7A-7B and 20). Curiously, the secretory subtype did not show significant association with survival possibly due to the uniformly high expression of immune cells in the secretory subtype preventing demonstration of an incremental survival benefit per unit increase. Overall, in SQ, only the primitive subtype demonstrated significant immune cell expression associations with improved survival (see FIGs. 7A-7B and 20).

## Conclusion

[00235] The 80 gene signature for SQ subtyping described in Example 2 showed similar results to the SQ subtyping gene signature(s) used in Example 1 in terms of showing how Lung SQ subtypes vary in their immune landscape. In agreement with the SQ subtyping gene signatures of Example 1, the SQ subtyping gene signature used in this example shows that Lung

SQ gene expression subtypes vary in their immune landscape. Intrinsic biologic subtypes of SQ reveal key differences in immune cell activation, which were not always correlated with *CD274* expression and demonstrated variable association with survival. SQ Classical subtype showed minimal immune infiltration suggesting reduced response to immunoRX. In SQ, subtype appeared to be a better predictor of immune infiltration than *CD274* (PD-L1). *CD274* expression was not associated with AIC expression nor with improved survival in SQ. The SQ primitive subtype showed immune feature expression associated with improved survival.

### **Incorporation by reference**

[00236] The following references are incorporated by reference in their entireties for all purposes.

[00237] 1.) Bindea et al., Immunity 2013; 39(4): 782-95. PMID 24138885

[00238] 2.) Wilkerson MD, et al. Clin Cancer Res 2010; 16(19): 4864-75. PMID 20643781

[00239] 3.) TCGA Lung SQCC. Nature 2012; 489(7417): 519-525. PMID 22960745

[00240] 4.) Lee ES, et al. Cancer Res 2008; 14(22): 7397-7404. PMID 19010856

[00241] 5.) Raponi M, et al. Cancer Res 2006; 66(7): 466-72. PMID 16885343

### **Example 4 – Expression Subtypes of Squamous Cell Carcinoma Reveal a Varied Immune Landscape and Unique Somatic Genetic Features Suggesting Differential Response to Multiple Drug Targets**

[00242] **Introduction:** Gene expression based subtyping in Lung Squamous Cell Carcinoma (SQ) classifies SQ tumors into distinct subtypes with variable outcomes and potential response to therapy. Gene expression based subtyping has consistently identified 4 subtypes with Lung SQ, Primitive, Classical, Basal and Secretory (1, 2) (see **FIG. 1**). SQ subtypes demonstrate key differences in genomic alterations, tumor drivers, prognosis, and likely response to various therapies (1-2).

[00243] **Methods:** As a follow up to the experiments conducted in Example 1, differential drug target gene expression was evaluated in the lung SQ subtypes from Example 1 that were determined using the TCGA lung cancer gene expression datasets (SQ n = 501)<sup>2</sup> shown in **FIG. 2**. Previously published SQ subtypes (Primitive, Classical, Secretory, or Basal) were defined in

Example 1 using gene expression patterns. In this example, the variable expression of genes from a clinical oncology solid tumor mutation panel (322 genes, see Table 8),<sup>3</sup> was examined in relation to SQ subtypes from Example 1 as a supplement to the examination of the immune cell gene signatures (Bindea et al. 24 immune cell types),<sup>4</sup> expression of single immune gene biomarkers (CTLA4, PDCD1 (PD-1), and CD274 (PD-L1)), proliferation (11 gene signature; see Table 9),<sup>5</sup> and non-silent mutation burden done in Example 1. Differential gene expression was assessed using the Kruskal-Wallis (KW) test with Bonferroni correction, while linear regression and Spearman correlations were used to evaluate association of non-silent mutation burden, tumor subtype, and CD274 (PD-L1) expression with immune cell expression.

**[00244] Results:** As shown in FIG. 24, variable expression of 208/322 tumor panel genes 244/322 (76%) in SQ subtypes were observed (KW Bonferroni threshold  $p < 0.000155$ ). Most drug target genes, including but not limited to SOX2, TGFBR2, SMO, CSF1R, PIK3CA, and HGF in SQ, exhibited strong differential expression across the subtypes ( $p < 1E-28$ ). Further, the top 25 genes from the 322 genes of the clinical oncology solid tumor mutation panel showing differential gene expression across the SQ subtypes can be seen in Table 10. Immune cell expression was also highly variable across subtypes (see FIG. 3). The SQCC secretory subtype demonstrated the greatest immune cell expression while the Classical subtype of SQ demonstrated low expression of immune cells (see FIG. 3). In SQ tumors, subtype was a better predictor of adaptive immune cell expression than CD274 (PD-L1) (median F-test p-value and adjusted R-squared were  $2.16e-24$  and  $0.20$  for subtype versus  $1.86e-09$  and  $0.07$  for CD274) (see FIG. 6). Non-silent mutation burden was not strongly correlated with immune cell expression (Spearman correlation =  $-0.08$  in SQ). Overall, as shown in FIG. 23, there were significant SQ subtype differences in proliferation, non-silent mutation burden, and key drug targets CD274 (PD-L1), PDCD1 (PD-1), and CTLA4. SQ subtypes demonstrated significant differences in many drug target tumor panel genes and in immune cell expression but did not demonstrate differences in mutation burden.

**[00245] Conclusion:** Molecular subtypes of lung SQ vary in expression of the majority of key drug target genes included in a clinical solid tumor sequencing panel. Molecular subtypes of lung SQ revealed differential expression of host immune response and immune targets.

Evaluation of subtypes as potential biomarkers for drug sensitivity should be investigated alone, and in combination with immune cell features and key mutation targets.

### **Incorporation by reference**

[00246] The following references are incorporated by reference in their entireties for all purposes.

[00247] 1.) Wilkerson MD, et al. Clin Cancer Res 2010; 16(19): 4864-75. PMID 20643781

[00248] 2.) TCGA Lung SQCC. Nature 2012; 489(7417): 519-525. PMID 22960745

[00249] 3.) Foundation Medicine Solid Tumor Mutation Panel accessed October 2014.

[00250] 4.) Bindea et al., Immunity 2013; 39(4): 782-95. PMID 24138885

[00251] 5.) Neilson TO, et al. Clin Cancer Res 2010; 16(21): 522-5232. PMID 20837693.

### **Example 5: Expression Subtypes of Lung Squamous Cell Carcinoma Reveal a Varied Immune Landscape and Unique Somatic Genetic Features Suggesting Differential Response to Multiple Drug Targets**

[00252] **Introduction:** Just like in Example 4, the purpose of this Example was to assess the differential expression of clinically important genes across previously defined gene expression subtypes of Squamous Cell Carcinoma (SQ). In contrast to Example 4 where the SQ gene expression based subtyping was performed using the TCGA lung cancer gene expression datasets (SQ n = 501)<sup>2</sup> as described in Example 1, gene expression based SQ subtyping in this Example was performed using the \ 80 gene sets described in Examples 2. Further, the clinically important genes were 322 genes (see Table 8) that constituted a clinical solid tumor mutation sequencing panel used in the management of oncology patients to identify genomic alterations impacting therapeutic management and/or to determine eligibility for targeted drug clinical trials. Just like in Example 4, differences in tumor proliferation were also assessed across the SQ subtypes using an 11 gene proliferation signature (see Table 9).

[00253] **Methods:** Using the TCGA lung cancer gene expression datasets (Squamous Cell Carcinoma (SQ) n = 501),<sup>1</sup> differential drug target gene expression was evaluated in lung SQ subtypes. Subtype was defined in in SQ using the Clanc80 SQ subtyper (see Example 2 and described herein) as previously described (nearest centroid prediction).<sup>3</sup> SQ subtypes Primitive,

Classical, Secretory, Basal were examined. Variable expression of genes from a clinical oncology solid tumor mutation panel (322 genes),<sup>4</sup> was examined in relation to SQ subtypes. Differential gene expression was assessed using the Kruskal-Wallis (KW) test with Bonferroni correction. Further, a proliferation score was calculated as the average expression ( $\log_2(\text{RSEM}+1)$ ) of available genes in the 11-gene PAM50 proliferation signature<sup>5</sup>. Subtype-proliferation association was tested using the Kruskal-Wallis test.

**[00254] Results:** Similar to **FIG. 32**, **FIG. 25** showed variable expression of 227/322 (70%) across the SQ subtypes were observed (KW Bonferroni threshold  $p < 0.000155$ ). Further, just like in **FIG. 23** in Example 4, there were significant SQ subtype differences in proliferation (see **FIG. 26**). Moreover, the top 25 genes from the 322 genes of the clinical oncology solid tumor mutation panel showing differential gene expression across the SQ subtypes seen in **Table 11** are very similar to those found in **Table 10**.

**[00255] Conclusion:** Just like in Example 4, molecular subtypes of lung SQ vary in expression of the majority of key drug target genes included in a clinical solid tumor sequencing panel. Molecular subtypes of lung SQ revealed differential expression of host immune response and immune targets.

### **Incorporation by reference**

**[00256]** The following references are incorporated by reference in their entireties for all purposes.

**[00257]** 1.) TCGA Lung SQCC. Nature 2012; 489(7417): 519-525. PMID 22960745

**[00258]** 2.) Wilkerson MD, et al. Clin Cancer Res 2010; 16(19):4864-75. PMID 20643781

**[00259]** 3.) Foundation Medicine Solid Tumor Mutation Panel accessed October 6, 2014.

**[00260]** 4.) Neilson TO, Parker JS, Leung S, et al. Clin Cancer Res 2010; 16(21): 5222-5232. PMID 20837693

**Table 8.** 322 genes of a clinical solid tumor mutation sequencing panel<sup>4</sup>

|                        |                    |            |       |        |        |        |             |           |
|------------------------|--------------------|------------|-------|--------|--------|--------|-------------|-----------|
| ABL1                   | C11orf30<br>(EMSY) | DDR2       | FGFR4 | IL7R   | MET    | PIK3CA | SDHD        | TSHR      |
| ABL2                   | CARD11             | DICER1     | FH    | INHBA  | MITF   | PIK3CB | SETD2       | U2AF1     |
| ACVR1<br>B             | CBFB               | DNMT3<br>A | FLCN  | INPP4B | MLH1   | PIK3CG | SF3B1       | VEGF<br>A |
| AKT1                   | CBL                | DOT1L      | FLT1  | IRF2   | MPL    | PIK3R1 | SLIT2       | VHL       |
| AKT2                   | CCND1              | EGFR       | FLT3  | IRF4   | MRE11A | PIK3R2 | SMAD2       | WISP3     |
| AKT3                   | CCND2              | EP300      | FLT4  | IRS2   | MSH2   | PLCG2  | SMAD3       | WT1       |
| ALK                    | CCND3              | EPHA3      | FOXL2 | JAK1   | MSH6   | PMS2   | SMAD4       | XPO1      |
| AMER1<br>(FAM12<br>3B) | CCNE1              | EPHA5      | FOXP1 | JAK2   | MTOR   | POLD1  | SMARC<br>A4 | ZBTB2     |

|        |       |       |        |               |              |         |         |        |
|--------|-------|-------|--------|---------------|--------------|---------|---------|--------|
| APC    | CD274 | EPHA7 | FRS2   | JAK3          | MUTYH        | POLE    | SMARCB1 | ZNF217 |
| AR     | CD79A | EPHB1 | FUBP1  | JUN           | MYC          | PPP2R1A | SMO     | ZNF703 |
| ARAF   | CD79B | ERBB2 | GABRA6 | KAT6A (MYST3) | MYCL (MYCL1) | PRDM1   | SNCAIP  | ETV4   |
| ARFRP1 | CDC73 | ERBB3 | GATA1  | KDM5A         | MYCN         | PREX2   | SOCS1   | ETV5   |
| ARID1A | CDH1  | ERBB4 | GATA2  | KDM5C         | MYD88        | PRKAR1A | SOX10   | ETV6   |
| ARID1B | CDK12 | ERG   | GATA3  | KDM6A         | NF1          | PRKCI   | SOX2    | ETV1   |
| ARID2  | CDK4  | ERF1  | GATA4  | KDR           | NF2          | PRKDC   | SOX9    | NFKB1A |
| ASXL1  | CDK6  | ESR1  | GATA6  | KEAP1         | NFE2L2       | PRSS8   | SPEN    |        |

|       |        |            |                        |                 |        |        |       |  |
|-------|--------|------------|------------------------|-----------------|--------|--------|-------|--|
| ATM   | CDK8   | EZH2       | GID4<br>(C17orf<br>39) | KEL             | NFKBIA | PTCH1  | SPOP  |  |
| ATR   | CDKN1A | FAM46<br>C | GLI1                   | KIT             | NKX2-1 | PTEN   | SPTA1 |  |
| ATRX  | CDKN1B | FANCA      | GNA11                  | KLHL6           | NOTCH1 | PTPN11 | SRC   |  |
| AURKA | CDKN2A | FANCC      | GNA13                  | KMT2A<br>(MLL)  | NOTCH2 | QKI    | STAG2 |  |
| AURKB | CDKN2B | FANCD<br>2 | GNAQ                   | KMT2C<br>(MLL3) | NOTCH3 | RAC1   | STAT3 |  |
| AXIN1 | CDKN2C | FANCE      | GNAS                   | KMT2D<br>(MLL2) | NPM1   | RAD50  | STAT4 |  |
| AXL   | CEBPA  | FANCF      | GPR124                 | KRAS            | NRAS   | RAD51  | STK11 |  |
| BAP1  | CHD2   | FANCG      | GRIN2<br>A             | LMO1            | NSD1   | RAF1   | SUFU  |  |

|            |        |       |              |        |       |        |                             |  |
|------------|--------|-------|--------------|--------|-------|--------|-----------------------------|--|
| BARD1      | CHD4   | FANCL | GRM3         | LRP1B  | NTRK1 | RANBP2 | SYK                         |  |
| BCL2       | CHEK1  | FAS   | GSK3B        | LYN    | NTRK2 | RARA   | TAF1                        |  |
| BCL2L1     | CHEK2  | FAT1  | H3F3A        | LZTR1  | NTRK3 | RB1    | TBX3                        |  |
| BCL2L2     | CIC    | FBXW7 | HGF          | MAGI2  | NUP93 | RBM10  | TERC                        |  |
| BCOR       | CREBBP | FGF10 | HNF1A        | MAP2K1 | PAK3  | RET    | TERT<br>(promote<br>r only) |  |
| BCORL<br>1 | CRKL   | FGF14 | HRAS         | MAP2K2 | PALB2 | RICTOR | TET2                        |  |
| BLM        | CRLF2  | FGF19 | HSD3B<br>1   | MAP2K4 | PARK2 | RNF43  | TGFBR<br>2                  |  |
| BRAF       | CSF1R  | FGF23 | HSP90A<br>A1 | MAP3K1 | PAX5  | ROS1   | TNFAIP<br>3                 |  |

|       |        |       |       |       |          |         |          |  |
|-------|--------|-------|-------|-------|----------|---------|----------|--|
| BRCA1 | CTCF   | FGF3  | IDH1  | MCL1  | PBRM1    | RPTOR   | TNFRSF14 |  |
| BRCA2 | CTNNA1 | FGF4  | IDH2  | MDM2  | PDCD1LG2 | RUNX1   | TOP1     |  |
| BRD4  | CTNNB1 | FGF6  | IGF1R | MDM4  | PDGFRA   | RUNX1T1 | TOP2A    |  |
| BRIP1 | CUL3   | FGFR1 | IGF2  | MED12 | PDGFRB   | SDHA    | TP53     |  |
| BTG1  | CYLD   | FGFR2 | IKBKE | MEF2B | PDK1     | SDHB    | TSC1     |  |
| BTK   | DAXX   | FGFR3 | IKZF1 | MEN1  | PIK3C2B  | SDHC    | TSC2     |  |

| Table 9. 11 gene proliferation gene signature |               |       |       |
|-----------------------------------------------|---------------|-------|-------|
| BIRC5                                         | CDCA1 (NUF2)  | MKI67 | TYMS  |
| CCNB1                                         | CEP55         | PTTG1 | UBE2C |
| CDC20                                         | KNTC2 (NDC80) | RRM2  |       |

**[00261] Table 10.** Top 25 differentiated genes of the 322 tumor panel<sup>3</sup> for the SQ expression subtypes as determined in Example 4.

| <b>SQ Genes</b> | <b>KW p value</b> |
|-----------------|-------------------|
| NTRK2           | 2.41E-55          |
| SOX2            | 1.64E-54          |
| NFE2L2          | 1.05E-49          |
| TGFBR2          | 4.97E-40          |
| SMO             | 7.91E-40          |
| KEAP1           | 1.14E-38          |
| GATA3           | 1.02E-37          |
| JAK1            | 7.23E-37          |
| JAK3            | 1.34E-36          |
| CSF1R           | 3.16E-36          |
| FOXP1           | 4.18E-35          |
| AXL             | 3.32E-34          |
| PTCH1           | 2.00E-33          |
| STAT4           | 2.61E-32          |
| TNFRSF14        | 6.17E-32          |
| ESR1            | 4.46E-31          |
| BTK             | 5.65E-31          |
| FLT4            | 1.13E-30          |
| IKZF1           | 9.11E-30          |
| PIK3CA          | 4.49E-29          |
| HGF             | 6.64E-29          |
| LRP1B           | 1.63E-28          |
| FANCC           | 2.57E-28          |
| PIK3CG          | 9.86E-28          |
| GATA6           | 5.03E-27          |

**[00262] Table 11.** Top 25 differentiated genes of the 322 tumor panel<sup>3</sup> for the SQ expression subtypes as determined in Example 5.

| <b>SQ Genes</b> | <b>KW p value</b> |
|-----------------|-------------------|
| NTRK2           | 7.84E-59          |
| SOX2            | 4.41E-58          |
| NFE2L2          | 1.26E-48          |
| KEAP1           | 1.83E-41          |
| SMO             | 2.52E-41          |
| GATA3           | 6.38E-35          |
| FOXP1           | 2.88E-34          |
| JAK1            | 3.30E-34          |
| PTCH1           | 3.79E-34          |
| PIK3CA          | 7.26E-33          |
| LRP1B           | 9.39E-32          |
| JAK3            | 2.75E-31          |
| ESR1            | 1.40E-29          |
| GNA13           | 2.31E-29          |
| ETV4            | 6.10E-29          |
| FANCC           | 1.44E-28          |
| PRKCI           | 2.06E-28          |
| ERRFI1          | 1.75E-26          |
| AXL             | 3.74E-26          |
| TNFRSF14        | 6.69E-26          |
| TGFBR2          | 8.42E-26          |
| EZH2            | 3.56E-25          |
| FLT4            | 1.70E-24          |
| CSF1R           | 2.28E-24          |
| FGFR2           | 1.76E-23          |

**[00263]** The various embodiments described above can be combined to provide further embodiments. All of the U.S. patents, U.S. patent application publications, U.S. patent application, foreign patents, foreign patent application and non-patent publications referred to in this specification and/or listed in the Application Data Sheet are incorporated herein by

reference, in their entirety. Aspects of the embodiments can be modified, if necessary to employ concepts of the various patents, application and publications to provide yet further embodiments.

**[00264]** These and other changes can be made to the embodiments in light of the above-detailed description. In general, in the following claims, the terms used should not be construed to limit the claims to the specific embodiments disclosed in the specification and the claims, but should be construed to include all possible embodiments along with the full scope of equivalents to which such claims are entitled. Accordingly, the claims are not limited by the disclosure.

**What is claimed is:**

1. A method for determining a squamous cell carcinoma (SQ) subtype of a lung tissue sample obtained from a patient, the method comprising detecting an expression level of at least one classifier biomarker of **Table 1**, wherein the detection of the expression level of the classifier biomarker specifically identifies a basal, classical, secretory or primitive SQ subtype.
2. The method of claim 1, wherein the method further comprises comparing the detected levels of expression of the at least one classifier biomarkers of **Table 1** to the expression of the at least one classifier biomarkers of **Table 1** in at least one sample training set(s), wherein the at least one sample training set comprises expression data of the at least one classifier biomarkers of **Table 1** from a reference SQ basal sample, expression data of the at least one classifier biomarkers of **Table 1** from a reference SQ classical sample, expression data of the at least one classifier biomarkers of **Table 1** from a reference SQ secretory sample, expression data of the at least one classifier biomarkers of **Table 1** from a reference SQ primitive sample or a combination thereof; and classifying the sample as basal, classical, secretory or primitive subtype based on the results of the comparing step.
3. The method of claim 2, wherein the comparing step comprises applying a statistical algorithm which comprises determining a correlation between the expression data obtained from the sample and the expression data from the at least one training set(s); and classifying the sample as a basal, classical, secretory or primitive subtype based on the results of the statistical algorithm.
4. The method of any of the above claims, wherein the expression level of the classifier biomarker is detected at the nucleic acid level.
5. The method of claim 4, wherein the nucleic acid level is RNA or cDNA.
6. The method claim 4 or 5, wherein the detecting an expression level comprises performing quantitative real time reverse transcriptase polymerase chain reaction (qRT-PCR), RNAseq, microarrays, gene chips, nCounter Gene Expression Assay, Serial Analysis of Gene Expression (SAGE), Rapid Analysis of Gene Expression (RAGE), nuclease

protection assays, Northern blotting, or any other equivalent gene expression detection techniques.

7. The method of claim 6, wherein the expression level is detected by performing qRT-PCR.
8. The method of claim 7, wherein the detection of the expression level comprises using at least one pair of oligonucleotide primers specific for at least one classifier biomarker of **Table 1**.
9. The method of any of the above claims, wherein the sample is a formalin-fixed, paraffin-embedded (FFPE) lung tissue sample, fresh or a frozen tissue sample, an exosome, wash fluids, cell pellets, or a bodily fluid obtained from the patient.
10. The method of claim 9, wherein the bodily fluid is blood or fractions thereof, urine, saliva, or sputum.
11. The method of any one of the above claims, wherein the at least one classifier biomarker comprises a plurality of classifier biomarkers.
12. The method of claim 11, wherein the plurality of classifier biomarkers comprises at least two classifier biomarkers, at least 10 classifier biomarkers, at least 20 classifier biomarkers, at least 30 classifier biomarkers, at least 40 classifier biomarkers, at least 50 classifier biomarkers, at least 60 classifier biomarkers, or at least 70 classifier biomarkers of **Table 1**.
13. The method of any of claims 1-10, wherein the at least one classifier biomarker comprises all the classifier biomarkers of **Table 1**.
14. A method for determining a squamous cell carcinoma (SQ) subtype of a lung tissue sample obtained from a patient comprising detecting an expression level of at least one nucleic acid molecule that encodes a classifier biomarker having a specific expression pattern in lung cancer cells, wherein the classifier biomarker is selected from the group consisting of the classifier genes set forth in **Table 1**, the method comprising: (a) isolating nucleic acid material from a lung tissue sample from a patient; (b) mixing the nucleic acid material with oligonucleotides that are substantially complementary to portions of nucleic acid molecule of the classifier biomarker; and (c) detecting expression of the classifier biomarker.

15. The method of claim 14, wherein the method further comprises comparing the detected levels of expression of the at least one classifier biomarkers of **Table 1** to the expression of the at least one classifier biomarkers of **Table 1** in at least one sample training set(s), wherein the at least one sample training set comprises expression data of the at least one classifier biomarkers of **Table 1** from a reference SQ basal sample, expression data of the at least one classifier biomarkers of **Table 1** from a reference SQ classical sample, expression data of the at least one classifier biomarkers of **Table 1** from a reference SQ secretory sample, expression data of the at least one classifier biomarkers of **Table 1** from a reference SQ primitive sample or a combination thereof; and classifying the sample as basal, classical, secretory or primitive subtype based on the results of the comparing step.
16. The method of claim 15, wherein the comparing step comprises applying a statistical algorithm which comprises determining a correlation between the expression data obtained from the sample and the expression data from the at least one training set(s); and classifying the sample as a basal, classical, secretory or primitive subtype based on the results of the statistical algorithm.
17. The method of any of claims 14-16, wherein the detecting the expression level comprises performing qRT-PCR or any hybridization-based gene assays.
18. The method of claim 17, wherein the expression level is detected by performing qRT-PCR.
19. The method of claim 18, wherein the detection of the expression level comprises using at least one pair of oligonucleotide primers specific for at least one classifier biomarker of **Table 1**.
20. The method of any of claims 14-19, further comprising predicting the response to a therapy for treating a subtype of lung squamous cell carcinoma (SQ) based on the detected expression level of the classifier biomarker.
21. The method of claim 20, wherein the therapy is chemotherapy, angiogenesis inhibitors and/or immunotherapy.
22. The method of claim 21, wherein the subtype of lung SQ is primitive and the therapy is an immunotherapy.

23. The method of any one of claims 14-22, wherein the sample is a formalin-fixed, paraffin-embedded (FFPE) lung tissue sample, fresh or a frozen tissue sample, an exosome, wash fluids, cell pellets or a bodily fluid obtained from the patient.
24. The method of claim 23, wherein the bodily fluid is blood or fractions thereof, urine, saliva, or sputum.
25. The method of any of claims 14-24, wherein the at least one nucleic acid molecule that encodes a classifier biomarker comprises a plurality of nucleic acid molecules that encode a plurality of classifier biomarkers.
26. The method of claim 25, wherein the plurality of classifier biomarkers comprises at least two classifier biomarkers, at least 10 classifier biomarkers, at least 20 classifier biomarkers, at least 30 classifier biomarkers, at least 40 classifier biomarkers, at least 50 classifier biomarkers, at least 60 classifier biomarkers, or at least 70 classifier biomarkers selected from **Table 1**.
27. The method of any of claims 14-24, wherein the at least one nucleic acid molecule that encodes a classifier biomarker comprises a plurality of nucleic acid molecules that include all the classifier biomarkers of **Table 1**.
28. A method of detecting a biomarker in a lung tissue sample obtained from a patient, the method comprising measuring the expression level of a plurality of biomarker nucleic acids selected from **Table 1** using an amplification, hybridization and/or sequencing assay.
29. The method of claim 28, wherein the lung tissue sample was previously diagnosed as being squamous cell carcinoma.
30. The method of claim 29, wherein the previous diagnosis was by histological examination.
31. The method of any of claims 28-30, wherein the amplification, hybridization and/or sequencing assay comprises performing quantitative real time reverse transcriptase polymerase chain reaction (qRT-PCR), RNAseq, microarrays, gene chips, nCounter Gene Expression Assay, Serial Analysis of Gene Expression (SAGE), Rapid Analysis of Gene Expression (RAGE), nuclease protection assays, Northern blotting, or any other equivalent gene expression detection techniques.
32. The method of claim 31, wherein the expression level is detected by performing qRT-PCR.

33. The method of claim 32, wherein the detection of the expression level comprises using at least one pair of oligonucleotide primers per each of the plurality of biomarker nucleic acids selected from **Table 1**.
34. The method of any of claims 28-33, wherein the sample is a formalin-fixed, paraffin-embedded (FFPE) lung tissue sample, fresh or a frozen tissue sample, an exosome, wash fluids, cell pellets, or a bodily fluid obtained from the patient.
35. The method of claim 34, wherein the bodily fluid is blood or fractions thereof, urine, saliva, or sputum.
36. The method of any of claims 28-35, wherein the plurality of biomarker nucleic acids comprises, consists essentially of or consists of at least two biomarker nucleic acids, at least 10 biomarker nucleic acids, at least 20 biomarker nucleic acids, at least 30 biomarker nucleic acids, at least 40 biomarker nucleic acids, at least 50 biomarker nucleic acids, at least 60 biomarker nucleic acids, or at least 70 biomarker nucleic acids of **Table 1**.
37. The method of any of claims 28-35, wherein the plurality of biomarker nucleic acids comprises, consists essentially of or consists of all the classifier biomarker nucleic acids of **Table 1**.
38. A method of detecting a biomarker in a lung tissue sample obtained from a patient, the method consisting essentially of measuring the expression level of a plurality of biomarker nucleic acids selected from **Table 1** using an amplification, hybridization and/or sequencing assay.
39. The method of claim 38, wherein the lung tissue sample was previously diagnosed as being squamous cell carcinoma.
40. The method of claim 39, wherein the previous diagnosis was by histological examination.
41. The method of any of claims 38-40, wherein the amplification, hybridization and/or sequencing assay comprises performing quantitative real time reverse transcriptase polymerase chain reaction (qRT-PCR), RNAseq, microarrays, gene chips, nCounter Gene Expression Assay, Serial Analysis of Gene Expression (SAGE), Rapid Analysis of Gene Expression (RAGE), nuclease protection assays, Northern blotting, or any other equivalent gene expression detection techniques.

42. The method of claim 41, wherein the expression level is detected by performing qRT-PCR.
43. The method of claim 42, wherein the detection of the expression level comprises using at least one pair of oligonucleotide primers per each of the plurality of biomarker nucleic acids selected from **Table 1**.
44. The method of any of claims 38-43, wherein the sample is a formalin-fixed, paraffin-embedded (FFPE) lung tissue sample, fresh or a frozen tissue sample, an exosome, wash fluids, cell pellets, or a bodily fluid obtained from the patient.
45. The method of claim 44, wherein the bodily fluid is blood or fractions thereof, urine, saliva, or sputum.
46. The method of any of claims 38-45, wherein the plurality of biomarker nucleic acids comprises, consists essentially of or consists of at least two biomarker nucleic acids, at least 10 biomarker nucleic acids, at least 20 biomarker nucleic acids, at least 30 biomarker nucleic acids, at least 40 biomarker nucleic acids, at least 50 biomarker nucleic acids, at least 60 biomarker nucleic acids, or at least 70 biomarker nucleic acids of **Table 1**.
47. The method of any of claims 38-45, wherein the plurality of biomarker nucleic acids comprises, consists essentially of or consists of all the classifier biomarker nucleic acids of **Table 1**.
48. A method of detecting a biomarker in a lung tissue sample obtained from a patient, the method consisting of measuring the expression level of a plurality of biomarker nucleic acids selected from **Table 1** using an amplification, hybridization and/or sequencing assay.
49. The method of claim 48, wherein the lung tissue sample was previously diagnosed as being squamous cell carcinoma.
50. The method of claim 49, wherein the previous diagnosis was by histological examination.
51. The method of any of claims 48-50, wherein the amplification, hybridization and/or sequencing assay comprises performing quantitative real time reverse transcriptase polymerase chain reaction (qRT-PCR), RNAseq, microarrays, gene chips, nCounter Gene Expression Assay, Serial Analysis of Gene Expression (SAGE), Rapid Analysis of Gene

Expression (RAGE), nuclease protection assays, Northern blotting, or any other equivalent gene expression detection techniques.

52. The method of claim 51, wherein the expression level is detected by performing qRT-PCR.
53. The method of claim 52, wherein the detection of the expression level comprises using at least one pair of oligonucleotide primers per each of the plurality of biomarker nucleic acids selected from **Table 1**.
54. The method of any of claims 48-53, wherein the sample is a formalin-fixed, paraffin-embedded (FFPE) lung tissue sample, fresh or a frozen tissue sample, an exosome, wash fluids, cell pellets, or a bodily fluid obtained from the patient.
55. The method of claim 54, wherein the bodily fluid is blood or fractions thereof, urine, saliva, or sputum.
56. The method of any of claims 48-55, wherein the plurality of biomarker nucleic acids comprises, consists essentially of or consists of at least two biomarker nucleic acids, at least 10 biomarker nucleic acids, at least 20 biomarker nucleic acids, at least 30 biomarker nucleic acids, at least 40 biomarker nucleic acids, at least 50 biomarker nucleic acids, at least 60 biomarker nucleic acids, or at least 70 biomarker nucleic acids of **Table 1**.
57. The method of any of claims 48-55, wherein the plurality of biomarker nucleic acids comprises, consists essentially of or consists of all the classifier biomarker nucleic acids of **Table 1**.
58. A method of determining whether a squamous cell carcinoma patient is likely to respond to immunotherapy, the method comprising,
  - determining the squamous cell carcinoma subtype of a lung tissue sample from the patient, wherein the squamous cell carcinoma subtype is selected from the group consisting of primitive, classical, secretory and basal; and
  - based on the subtype, assessing whether the patient is likely to respond to immunotherapy.
59. A method for selecting a squamous cell carcinoma patient for immunotherapy, the method comprising, determining a squamous cell carcinoma subtype of a lung tissue

- sample from the patient, based on the subtype; and selecting the patient for immunotherapy.
60. The method of claim 58 or 59, wherein the immunotherapy comprises checkpoint inhibitor therapy.
61. The method of claim 60, wherein the checkpoint inhibitor targets PD-1 or PD-L1.
62. The method of claim 60, wherein the checkpoint inhibitor targets CTLA-4.
63. The method of claim 61, wherein the checkpoint inhibitor is Pembrolizumab, Nivolumab or an antigen fragment binding fragment thereof.
64. The method of claim 62, wherein the checkpoint inhibitor is Ipilimumab or an antigen binding fragment thereof.
65. The method of any one of claims 58-64, wherein the patient is initially determined to have squamous cell carcinoma via a histological analysis of a sample.
66. The method of any one of claims 58-65, wherein the patient's squamous cell carcinoma subtype is selected from primitive, classical, secretory or basal, and is determined via a histological analysis of a sample obtained from the patient.
67. The method of any one of claims 65-66, wherein the sample is a formalin-fixed, paraffin-embedded (FFPE) lung tissue sample, fresh or a frozen tissue sample, an exosome, or a bodily fluid obtained from the patient.
68. The method of claim 67, wherein the bodily fluid is blood or fractions thereof, urine, saliva, or sputum.
69. The method of any one of claims 58-68, wherein the determining the squamous cell carcinoma subtype comprises determining expression levels of a plurality of classifier biomarkers.
70. The method of claim 69, wherein the determining the expression levels of the plurality of classifier biomarkers is at a nucleic acid level by performing RNA sequencing, reverse transcriptase polymerase chain reaction (RT-PCR) or hybridization based analyses.
71. The method of claim 69 or 70, wherein the plurality of classifier biomarkers for determining the squamous cell carcinoma subtype is selected from a publically available lung squamous cell carcinoma dataset.
72. The method of claim 71, wherein the publically available lung squamous cell carcinoma dataset is TCGA Lung SQ RNAseq dataset.

73. The method of claim 69, wherein the plurality of classifier biomarkers for determining the squamous cell carcinoma subtype is selected from **Table 1**.
74. The method of claim 70, wherein the RT-PCR is quantitative real time reverse transcriptase polymerase chain reaction (qRT-PCR).
75. The method of claim 74, wherein the RT-PCR is performed with primers specific to the plurality of classifier biomarkers of **Table 1**.
76. The method of any one of claims 69-75, further comprising comparing the detected levels of expression of the plurality of classifier biomarkers of **Table 1** to the expression of the plurality of classifier biomarkers of **Table 1** in at least one sample training set(s), wherein the at least one sample training set comprises expression data of the plurality of classifier biomarkers of **Table 1** from a reference squamous cell carcinoma basal sample, expression data of the plurality of classifier biomarkers of **Table 1** from a reference squamous cell carcinoma classical sample, expression data of the plurality of classifier biomarkers of **Table 1** from a reference squamous cell carcinoma primitive sample, expression data of the plurality of classifier biomarkers of **Table 1** from a reference squamous cell carcinoma secretory sample or a combination thereof; and classifying the first sample as basal, classical, primitive or secretory based on the results of the comparing step.
77. The method of claim 76, wherein the comparing step comprises applying a statistical algorithm which comprises determining a correlation between the expression data obtained from the sample and the expression data from the at least one training set(s); and classifying the sample as a basal, classical, primitive or secretory subtype based on the results of the statistical algorithm.
78. The method of any of claims 69-77, wherein the plurality of the classifier biomarkers comprise each of the classifier biomarkers set forth in **Table 1**.
79. A method of treating lung cancer in a subject, the method comprising:  
measuring the expression level of at least one biomarker nucleic acid in a lung cancer sample obtained from the subject, wherein the at least one biomarker nucleic acid is selected from a set of biomarkers listed in **Table 1**, wherein the presence, absence and/or level of the at least one biomarker indicates a subtype of the lung cancer; and administering an immunotherapeutic agent based on the subtype of the lung cancer.

80. The method of claim 79, wherein the lung cancer sample is a squamous cell carcinoma sample, and wherein the set of biomarkers is **Table 1**.
81. The method of claim 79, wherein the at least one biomarker nucleic acid selected from the set of biomarkers comprises, consists essentially of or consists of at least two biomarker nucleic acids, at least 10 biomarker nucleic acids, at least 20 biomarker nucleic acids, at least 30 biomarker nucleic acids, at least 40 biomarker nucleic acids, at least 50 biomarker nucleic acids, at least 60 biomarker nucleic acids, at least 70 biomarker nucleic acids or all of the biomarker nucleic acids of **Table 1**.
82. The method of any of claims 79-81, wherein the lung tissue sample was previously diagnosed as being squamous cell carcinoma.
83. The method of claim 82, wherein the previous diagnosis was by histological examination.
84. The method of any one of claims 79-83, further comprising measuring the expression of at least one biomarker from an additional set of biomarkers.
85. The method of claim 84, wherein the additional set of biomarkers comprise gene expression signatures of Innate Immune Cells (IIC), Adaptive Immune Cells (AIC), one or more individual immune biomarkers, one or more interferon (IFN) genes, one or more major histocompatibility complex, class II (MHCII) genes or a combination thereof.
86. The method of claim 84 or 85, wherein the additional set of biomarkers comprises genes selected from **Tables 4A, 4B, 5, 6, 7**, or a combination thereof.
87. The method of claim 86, wherein the gene expression signatures of AICs are selected from **Table 4A**.
88. The method of claim 86, wherein the gene expression signature of IICs are selected from **Table 4B**.
89. The method of claim 86, wherein the one or more individual immune biomarkers are selected from **Table 5**.
90. The method of claim 86, wherein the one or more IFN genes are selected from **Table 6**.
91. The method of claim 86, wherein the one or more MHCII genes are selected from **Table 7**.
92. The method of any of claims 79-91, wherein the measuring the expression level is conducted using an amplification, hybridization and/or sequencing assay.

93. The method of claim 92, wherein the amplification, hybridization and/or sequencing assay comprises performing quantitative real time reverse transcriptase polymerase chain reaction (qRT-PCR), RNAseq, microarrays, gene chips, nCounter Gene Expression Assay, Serial Analysis of Gene Expression (SAGE), Rapid Analysis of Gene Expression (RAGE), nuclease protection assays, Northern blotting, or any other equivalent gene expression detection techniques.
94. The method of claim 93, wherein the expression level is detected by performing qRT-PCR.
95. The method of any of claims 79-94, wherein the sample is a formalin-fixed, paraffin-embedded (FFPE) lung tissue sample, fresh or a frozen tissue sample, an exosome, wash fluids, cell pellets, or a bodily fluid obtained from the patient.
96. The method of claim 95, wherein the bodily fluid is blood or fractions thereof, urine, saliva, or sputum.
97. The method of any one of claims 79-96, wherein the subject's squamous cell carcinoma subtype is selected from primitive, classical, secretory or basal.
98. The method of claim 97, wherein the lung cancer subtype is primitive and wherein the immunotherapeutic agent comprises a checkpoint inhibitor.
99. The method of claim 98, wherein the checkpoint inhibitor targets PD-1 or PD-L1.
100. The method of claim 98, wherein the checkpoint inhibitor targets CTLA-4.
101. The method of claim 99, wherein the checkpoint inhibitor is Pembrolizumab, Nivolumab or an antigen fragment binding fragment thereof.
102. The method of claim 100, wherein the checkpoint inhibitor is Ipilimumab or an antigen binding fragment thereof.
103. The method of claim 79, wherein the at least one biomarker nucleic acid is a plurality of biomarker nucleic acids, wherein the plurality of biomarker nucleic acids comprises at least one biomarker nucleic acid listed in **Table 1** in combination with one or more biomarker nucleic acids from a publically available lung squamous cell carcinoma dataset, wherein the presence, absence and/or level of the plurality of biomarker nucleic acids indicates a subtype of the lung cancer.
104. The method of claim 79, wherein the at least one biomarker nucleic acid is a plurality of biomarker nucleic acids, wherein the plurality of biomarker nucleic acids

comprises all of the biomarker nucleic acids listed in **Table 1** in combination with one or more biomarker nucleic acids from a publically available lung squamous cell carcinoma dataset, wherein the presence, absence and/or level of the plurality of biomarker nucleic acids indicates a subtype of the lung cancer.

105. The method of claim 103 or 104, wherein the publically available lung squamous cell carcinoma dataset is TCGA Lung SQ RNAseq dataset.

FIG. 1  
Lung Cancer Subtyping

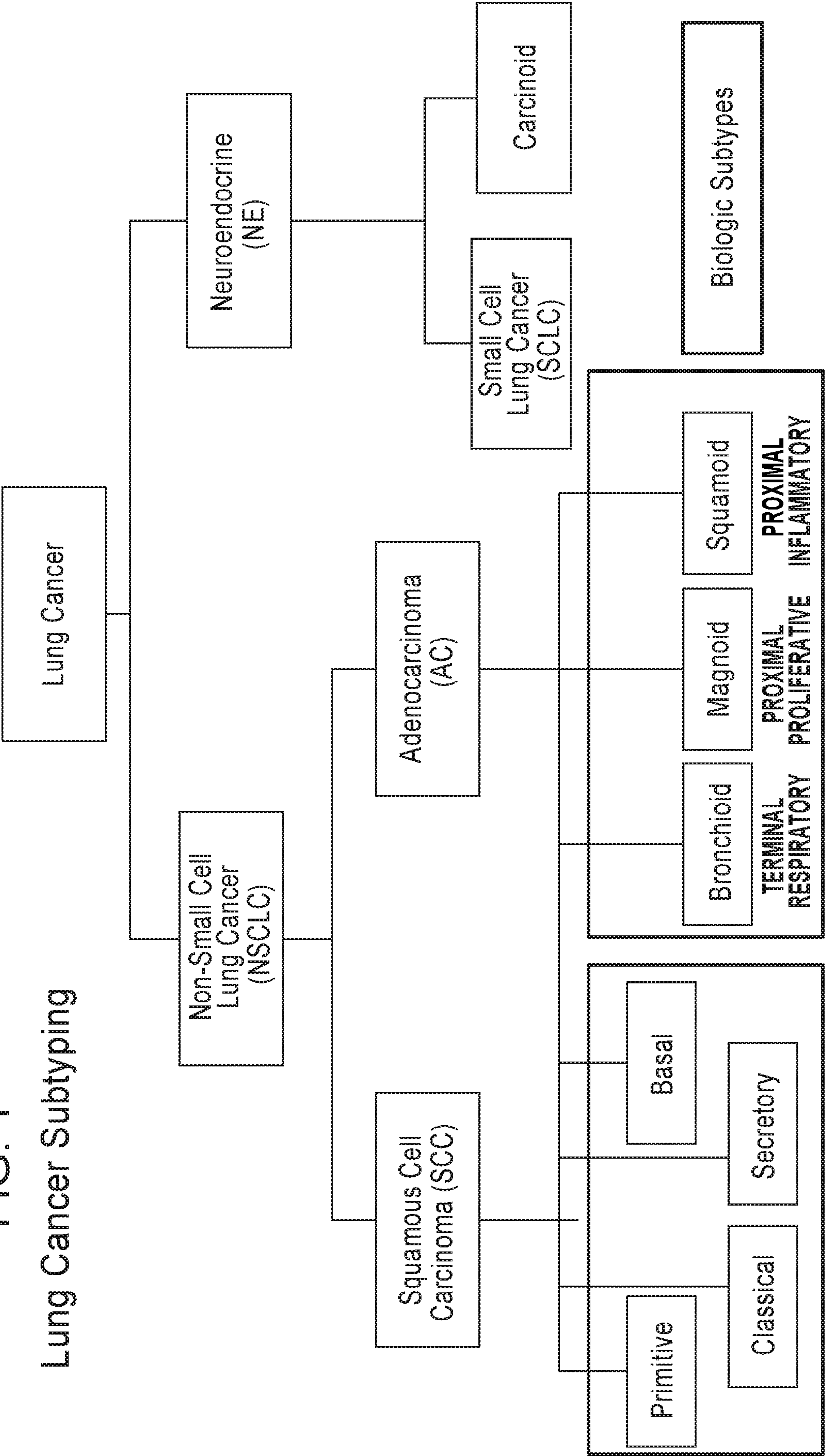


FIG. 2

| Characteristic               | TCGA <sup>4</sup> | Lee <sup>8</sup> | Raponi <sup>9</sup> | UNC <sup>3</sup> |
|------------------------------|-------------------|------------------|---------------------|------------------|
| Total # of samples           | 501               | 75               | 129                 | 56               |
| Tissue preservation          | Fresh Frozen      | Fresh Frozen     | Fresh Frozen        | Fresh Frozen     |
| Subtype                      |                   |                  |                     |                  |
| basal                        | 137               | 13               | 35                  | 12               |
| classical                    | 170               | 31               | 42                  | 21               |
| primitive                    | 73                | 12               | 23                  | 9                |
| secretory                    | 121               | 19               | 29                  | 14               |
| Gender                       |                   |                  |                     |                  |
| Female/Male/NA               | 125/363/13        | 5/70/0           | 47/82/0             | 24/32/0          |
| Age of diagnosis             |                   |                  |                     |                  |
| Median/(Range)               | 68/(39-90)        | 65/(41-82)       | 68/(42-91)          | 67/(41-85)       |
| Age not available            | 22                | 0                | 0                   | 0                |
| Stage                        |                   |                  |                     |                  |
| I                            | 241               | 0                | 73                  | 34               |
| II                           | 152               | 0                | 33                  | 19               |
| III                          | 85                | 0                | 23                  | 3                |
| IV                           | 7                 | 0                | 0                   | 0                |
| Stage not available          | 16                | 75               | 0                   | 0                |
| Smoking ever                 |                   |                  |                     |                  |
| yes                          | 416               | 0                | 119                 | 56               |
| no                           | 0                 | 0                | 4                   | 0                |
| Smoking status not available | 85                | 75               | 6                   | 0                |



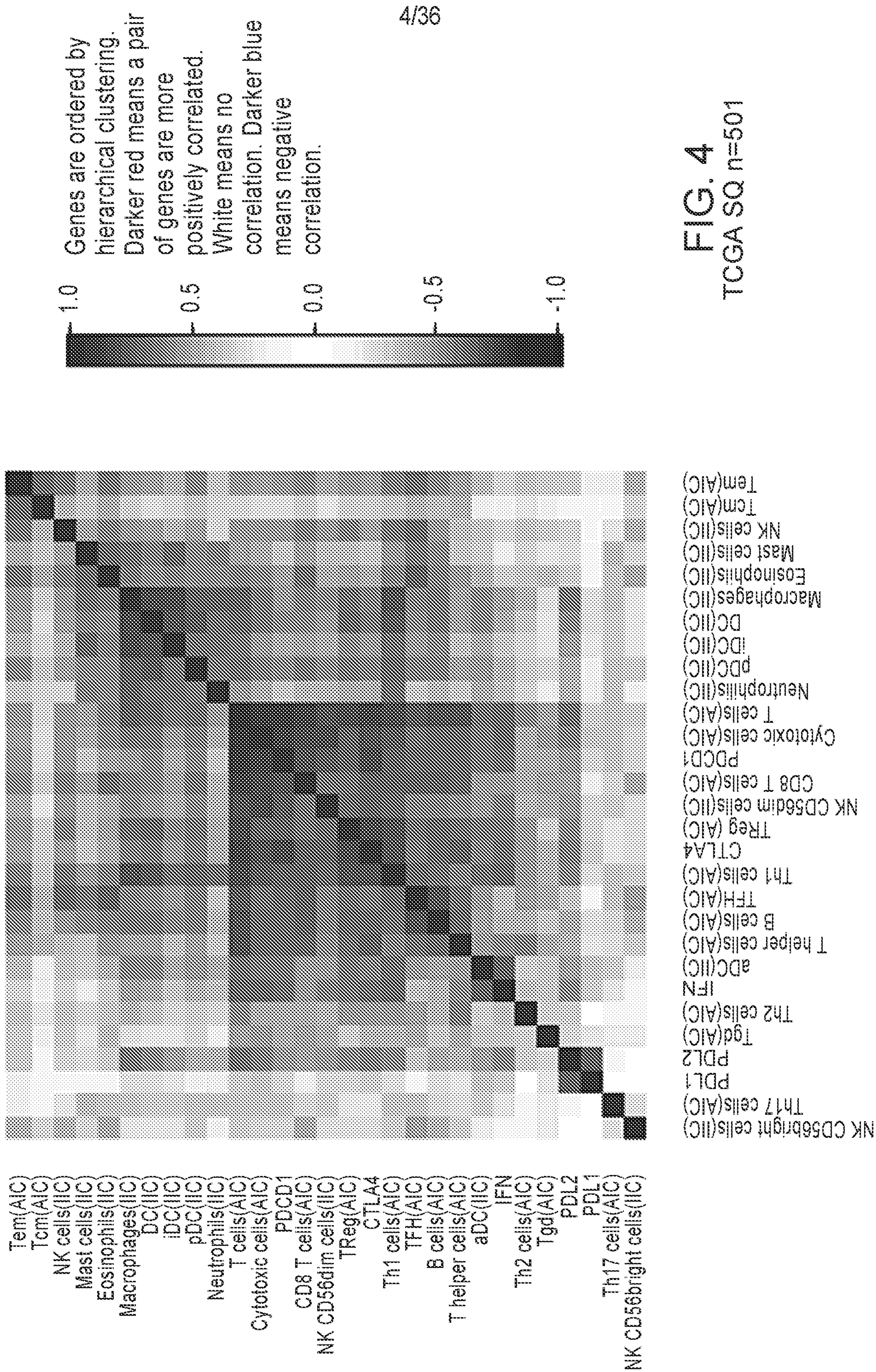


FIG. 4  
TCGA SQ n=501

FIG. 5

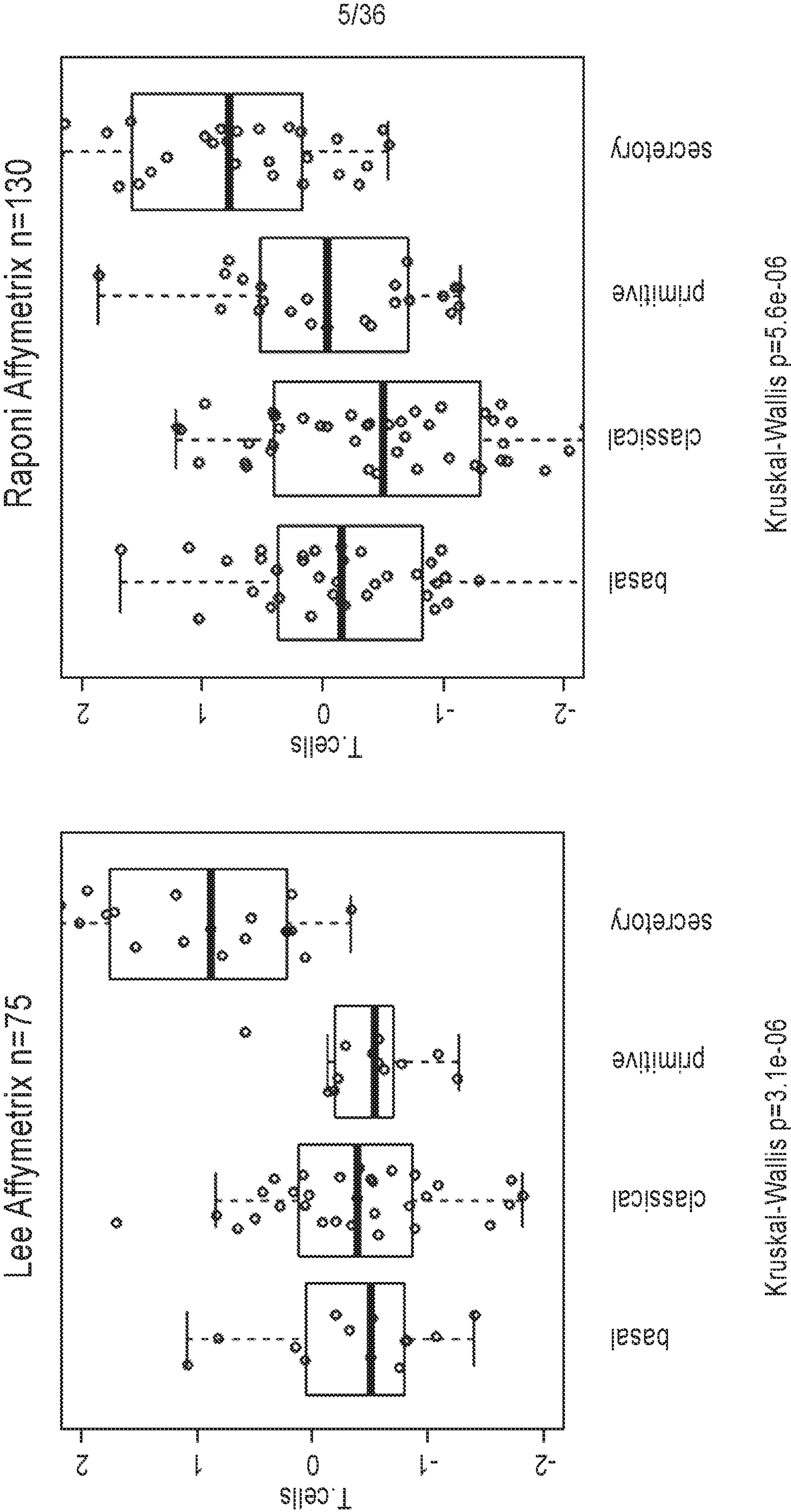


FIG. 5  
(continued)

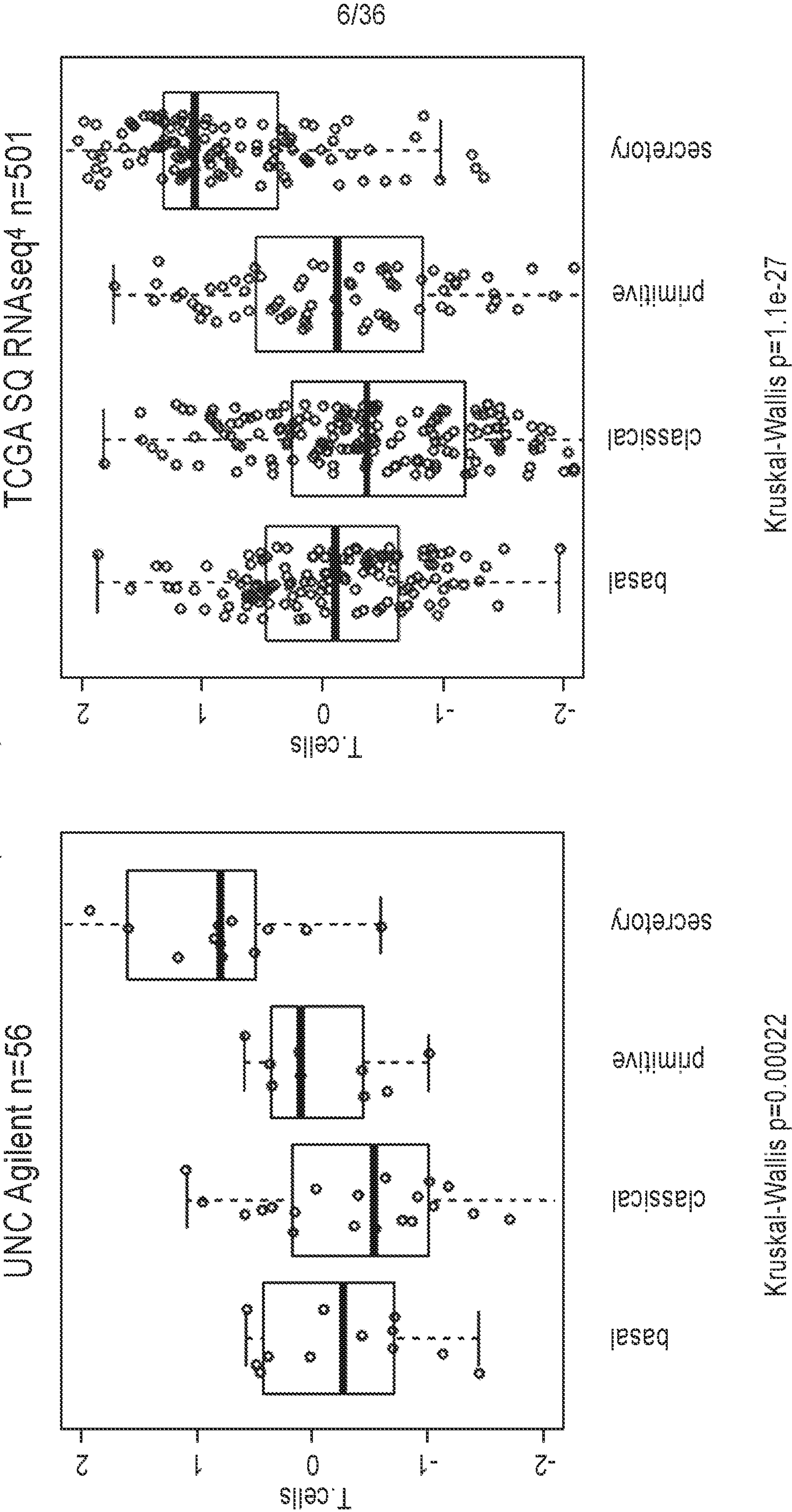


FIG. 6

TCGA SQ n=501

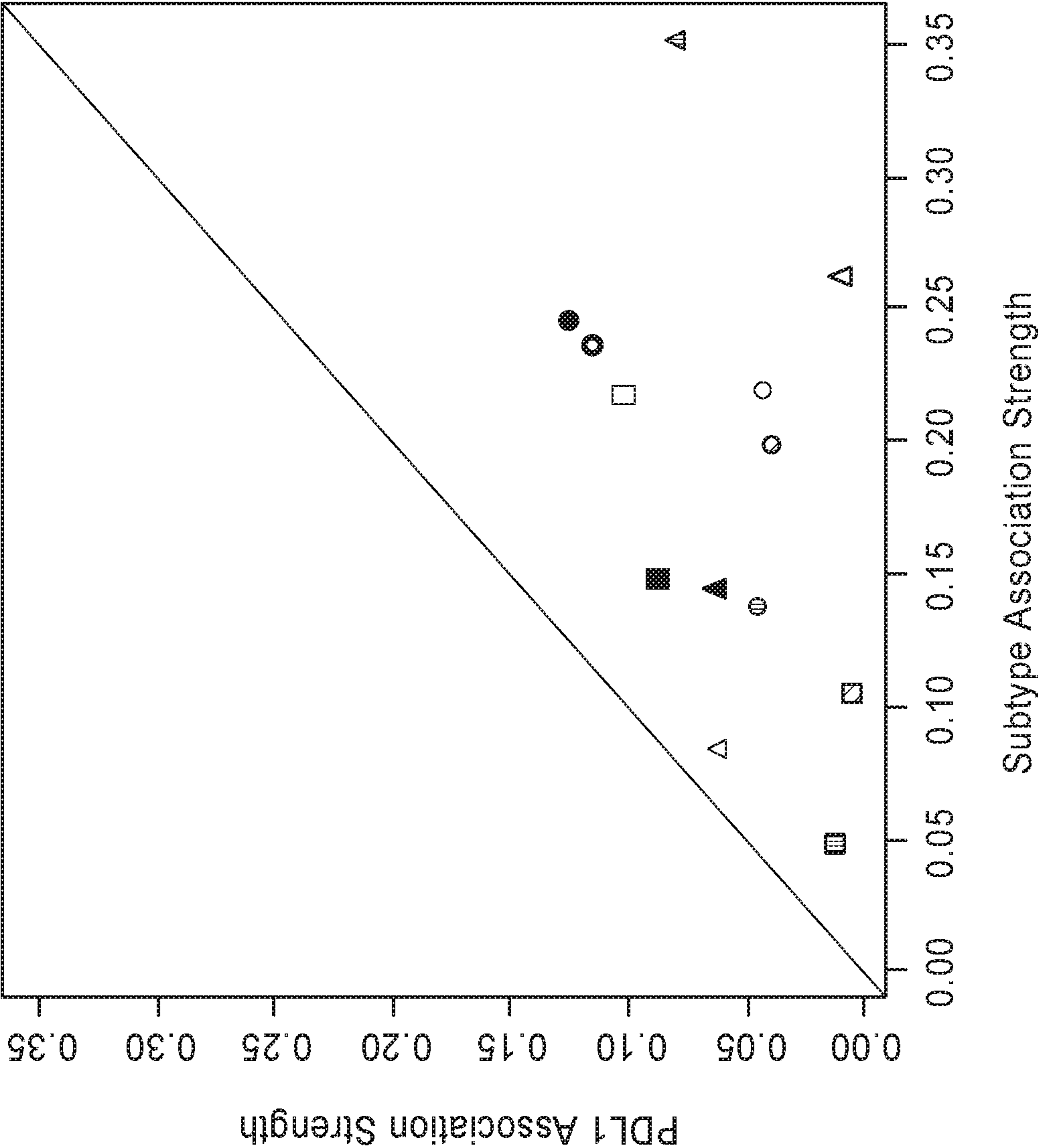
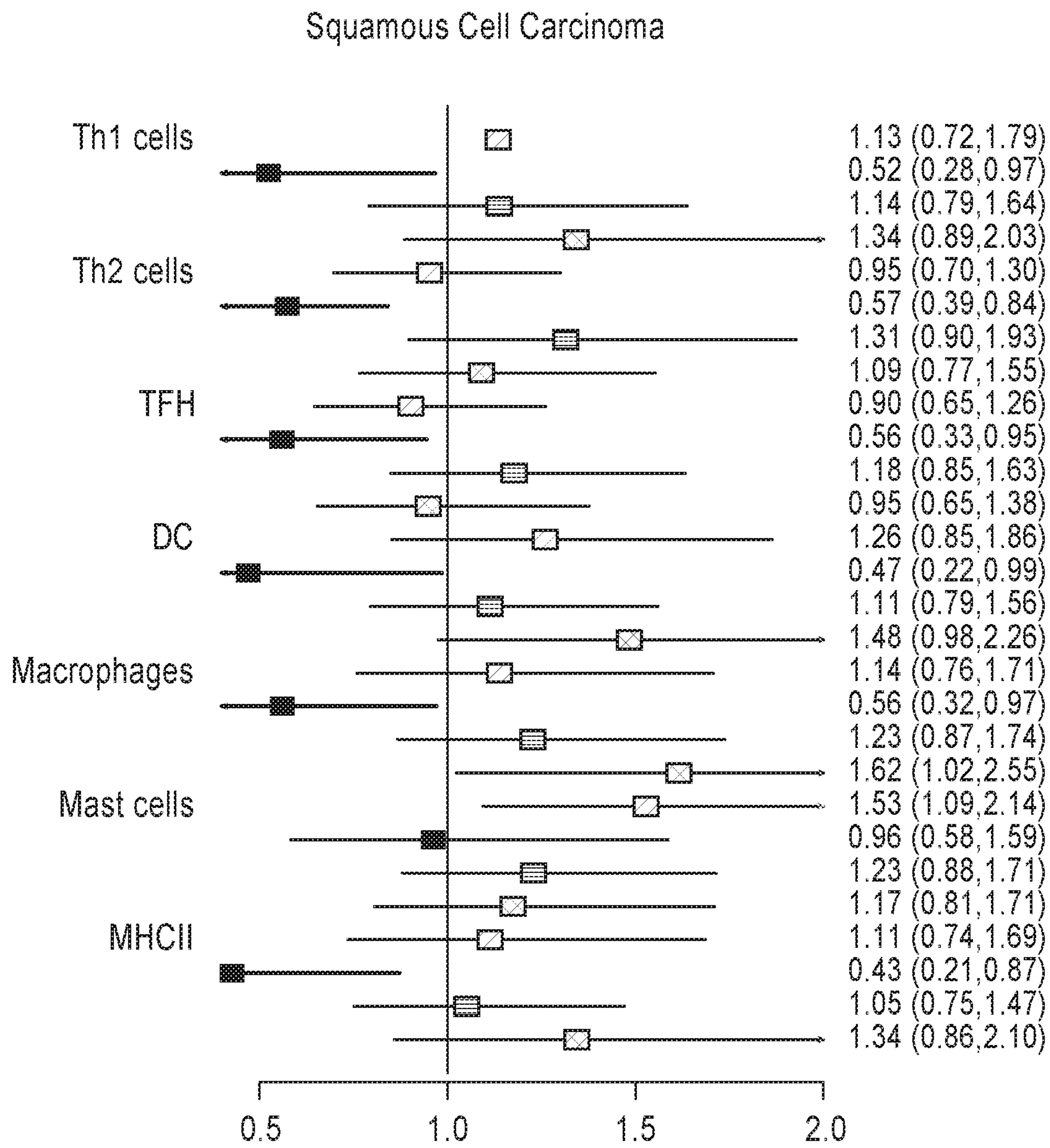


FIG. 7A



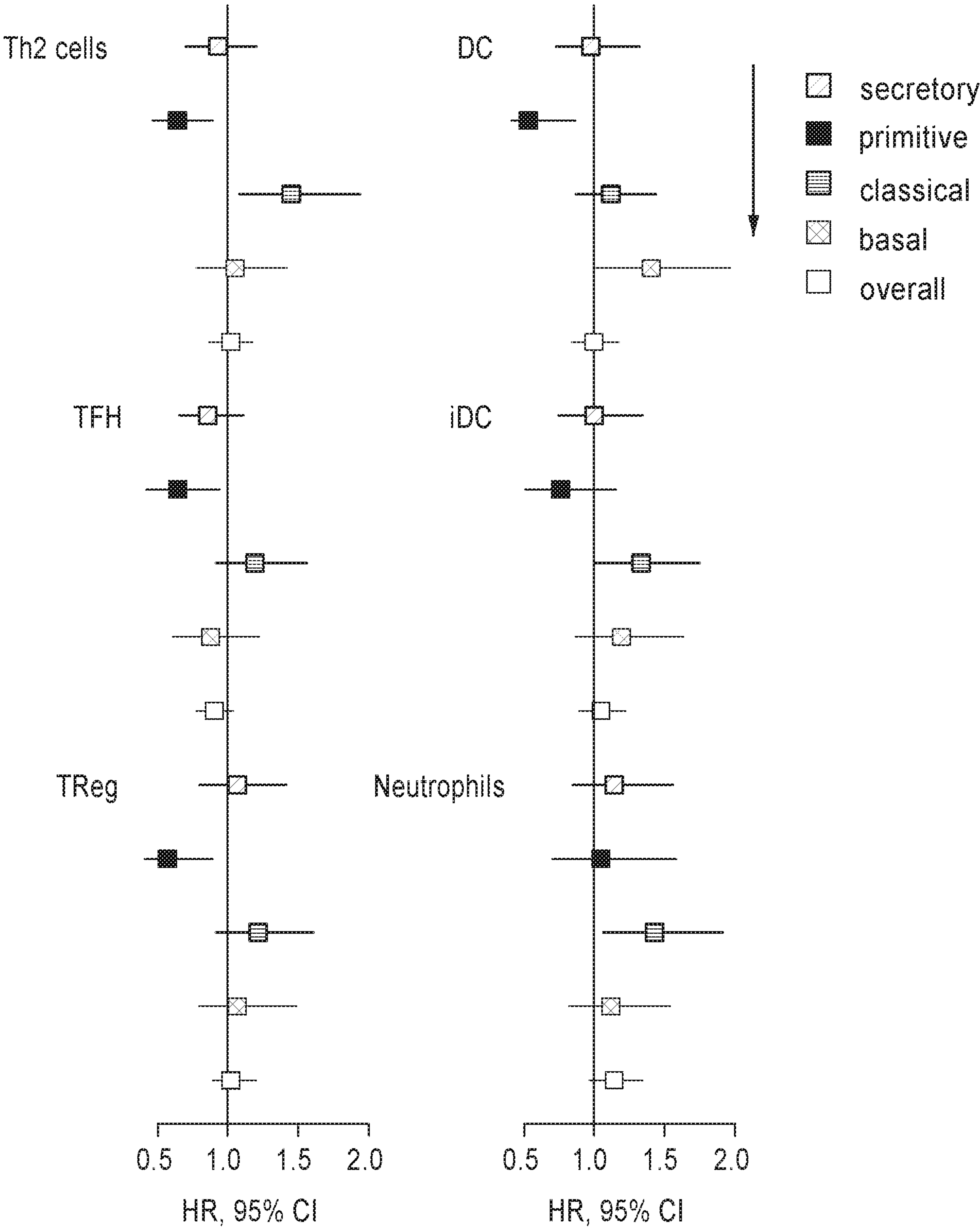
HR, 95% CI

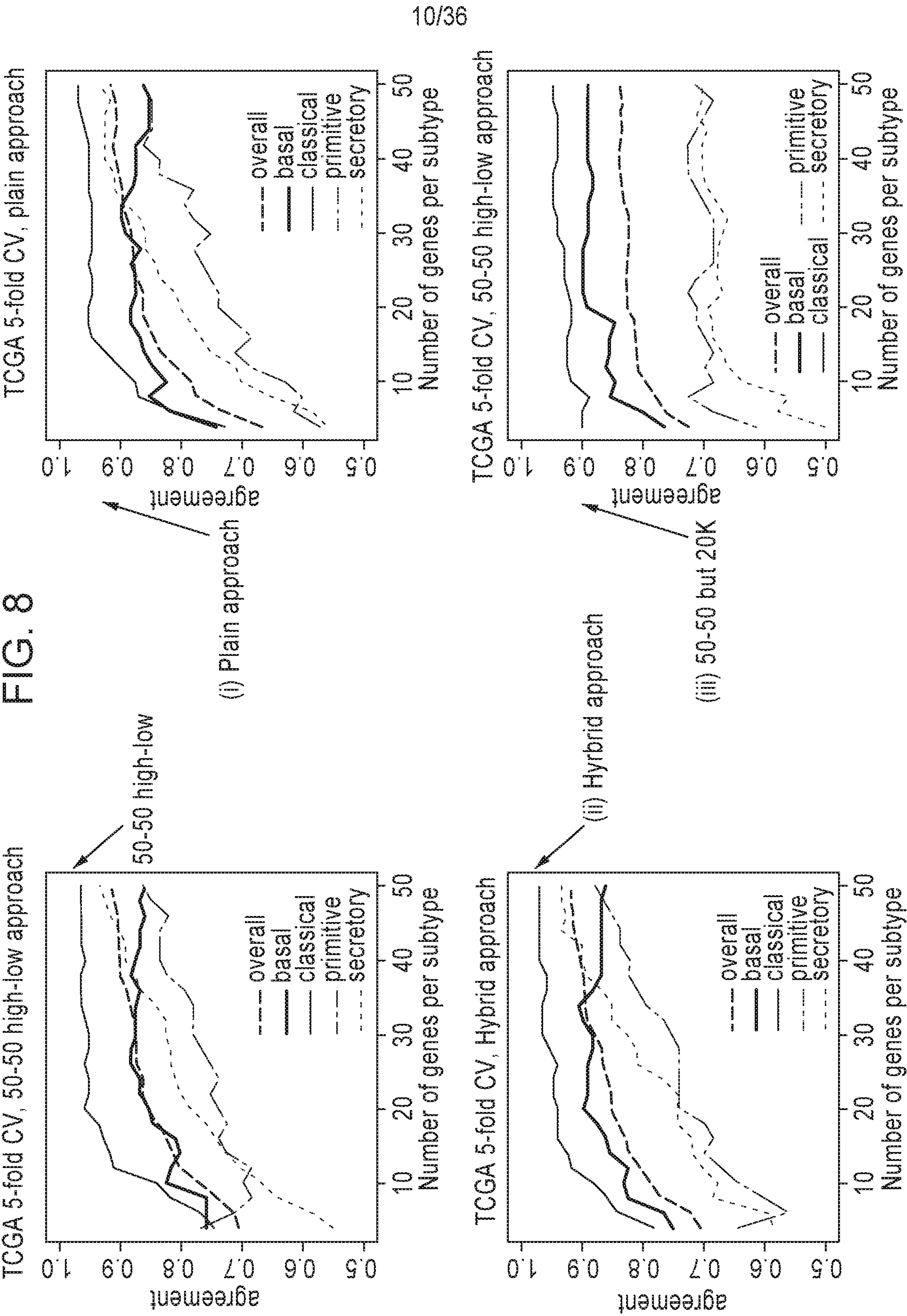
↓

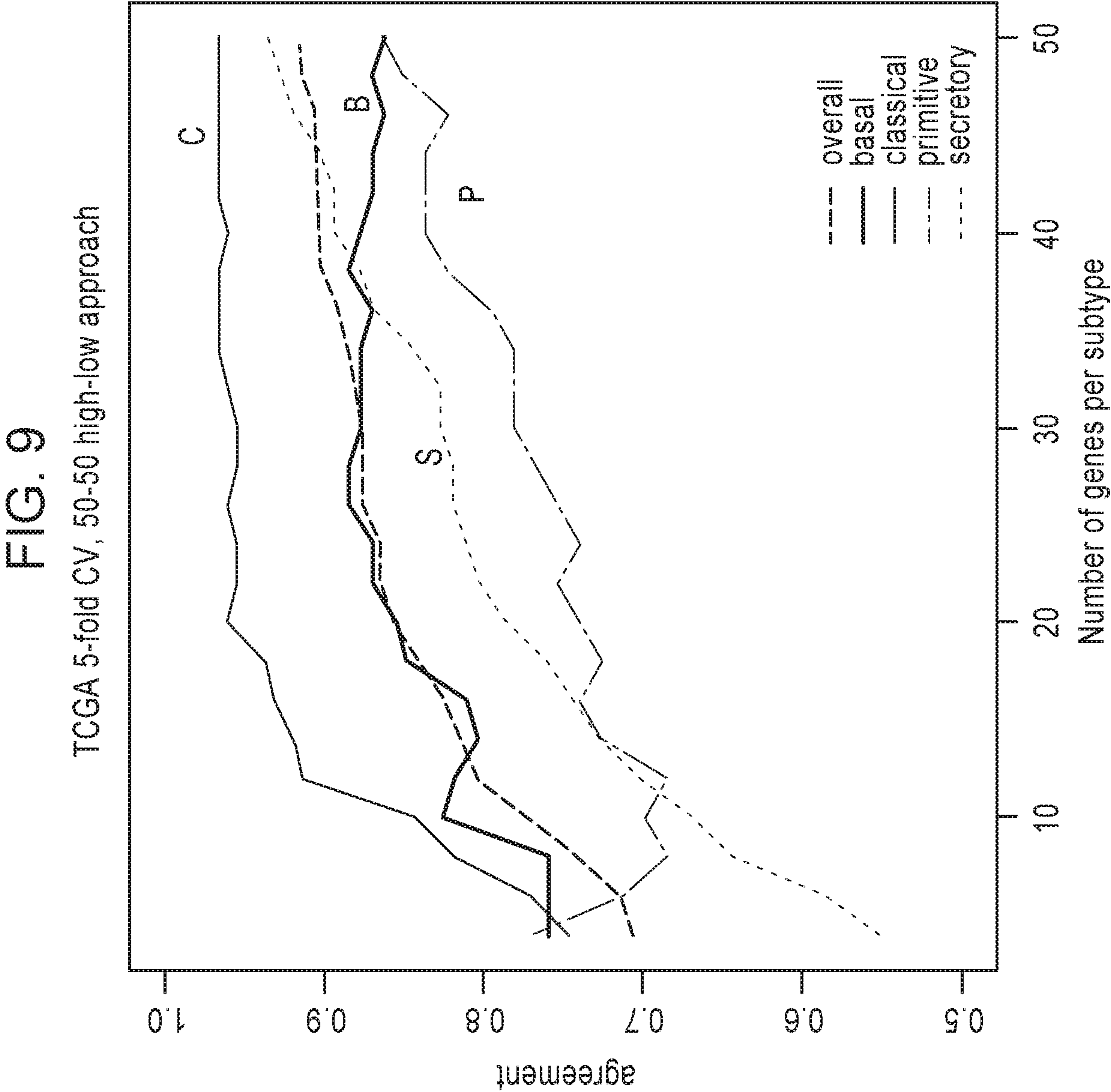
basal  
classical  
primitive  
secretory

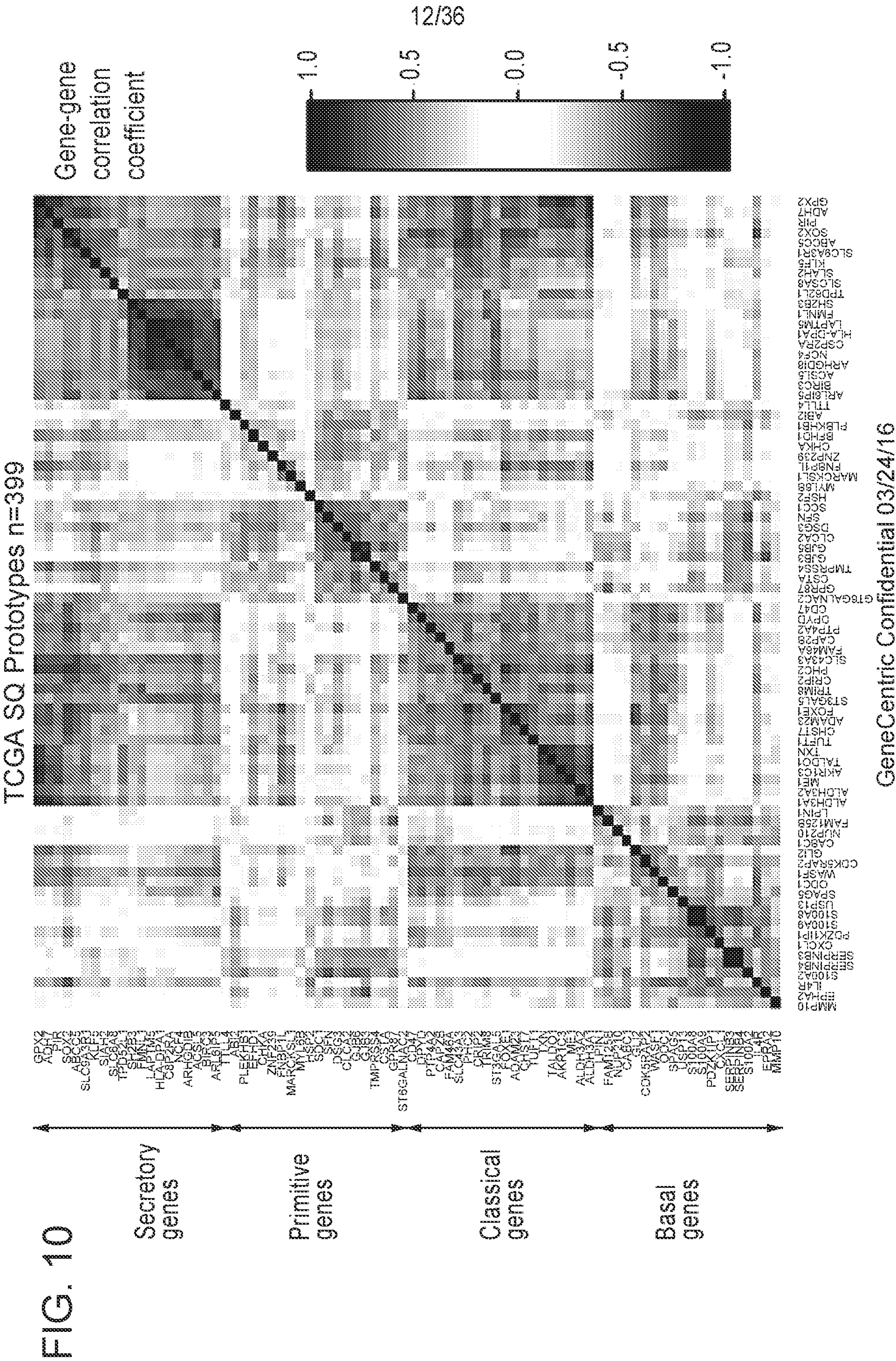
FIG. 7B

Squamous Cell Carcinoma (TCGA, UNC, Raponi)



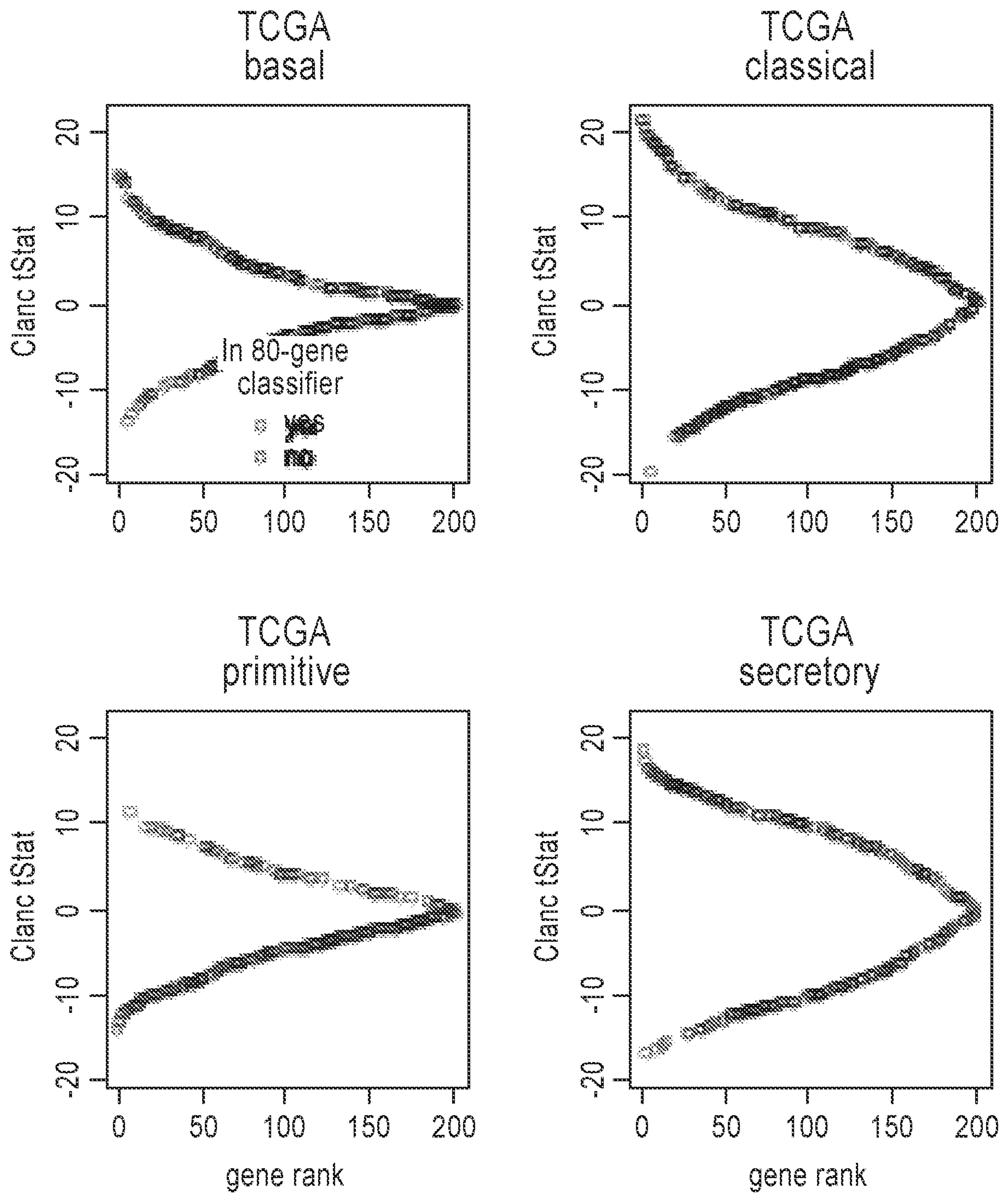






13/36

FIG. 11



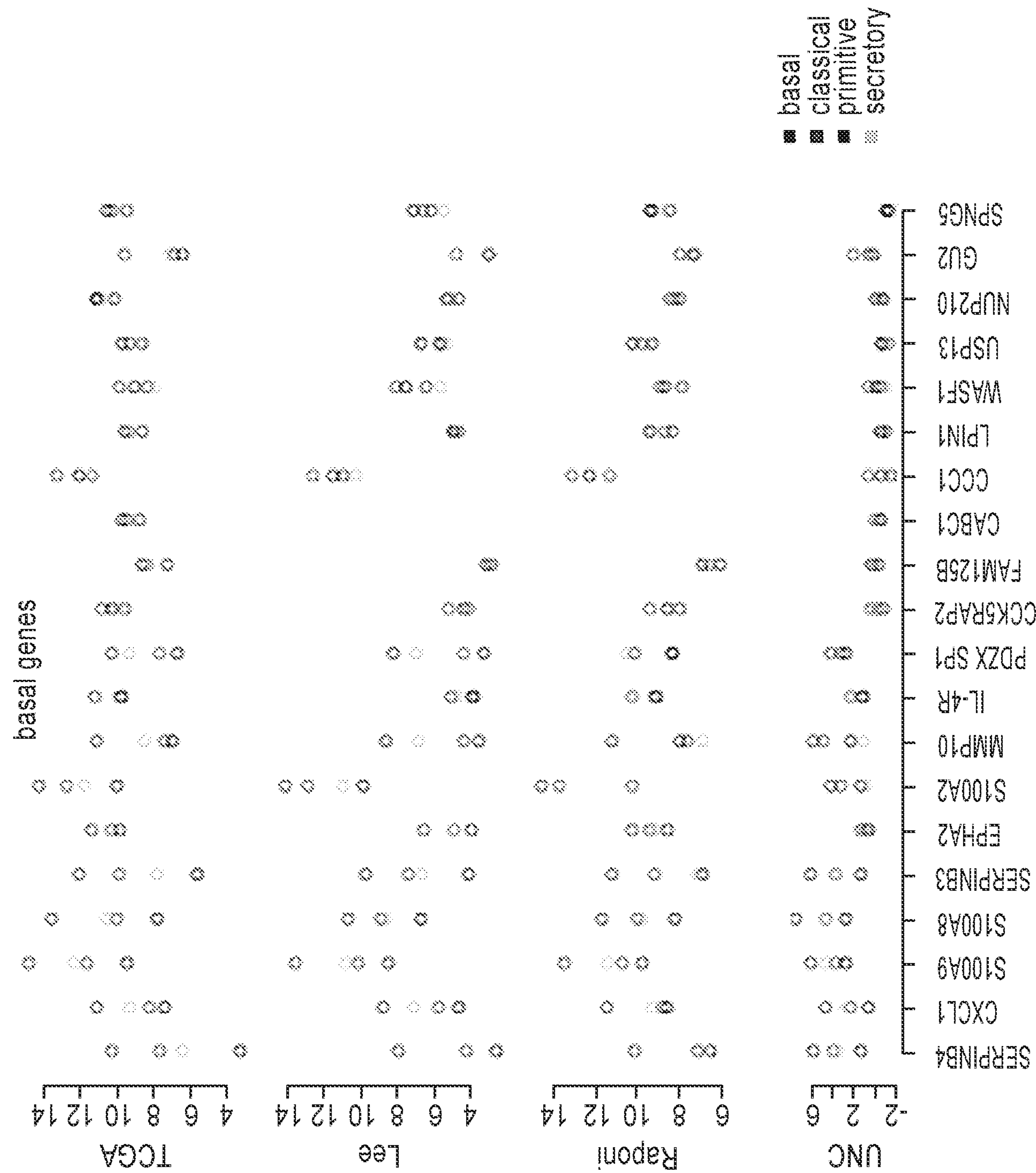


FIG. 12

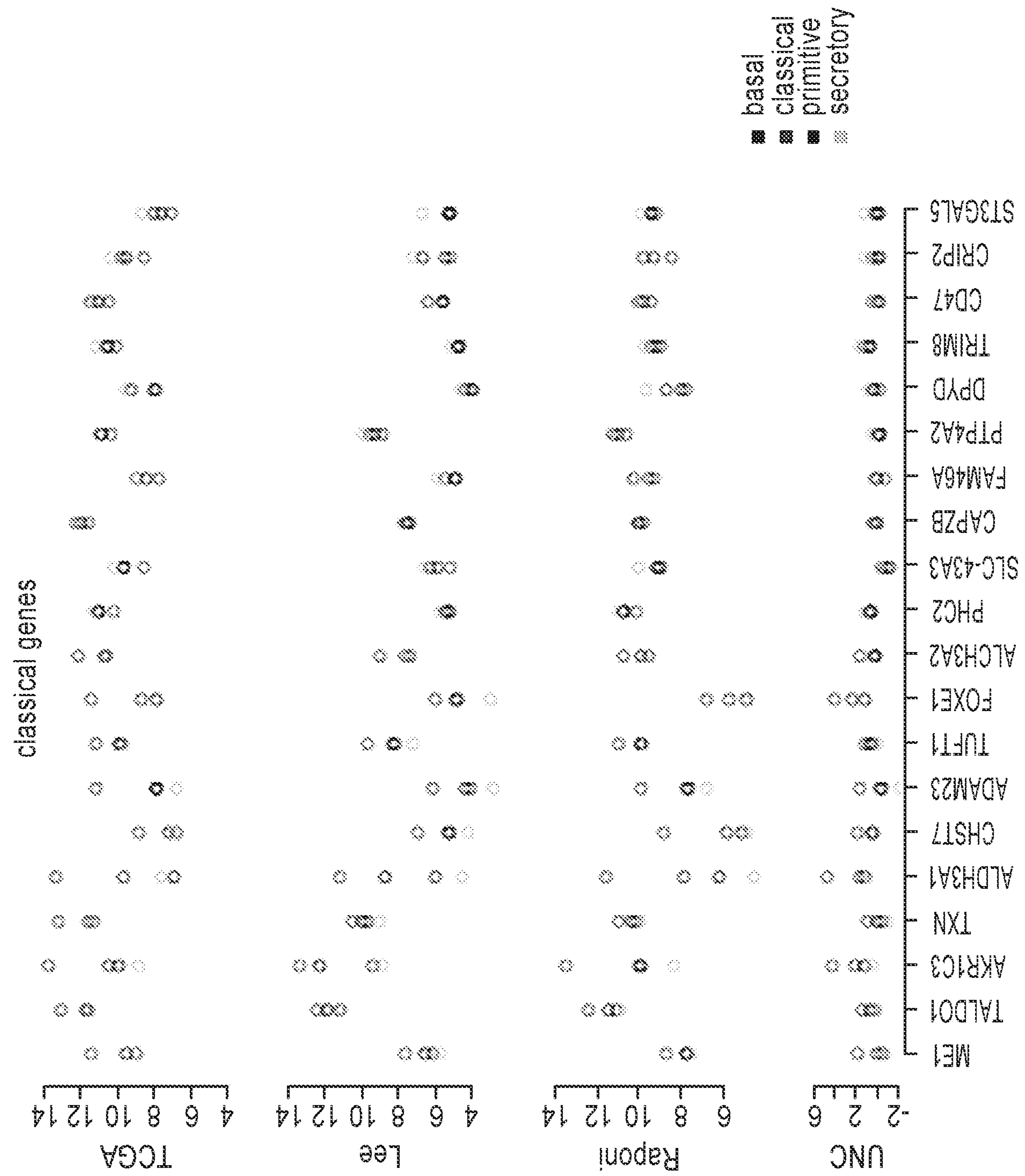


FIG. 13

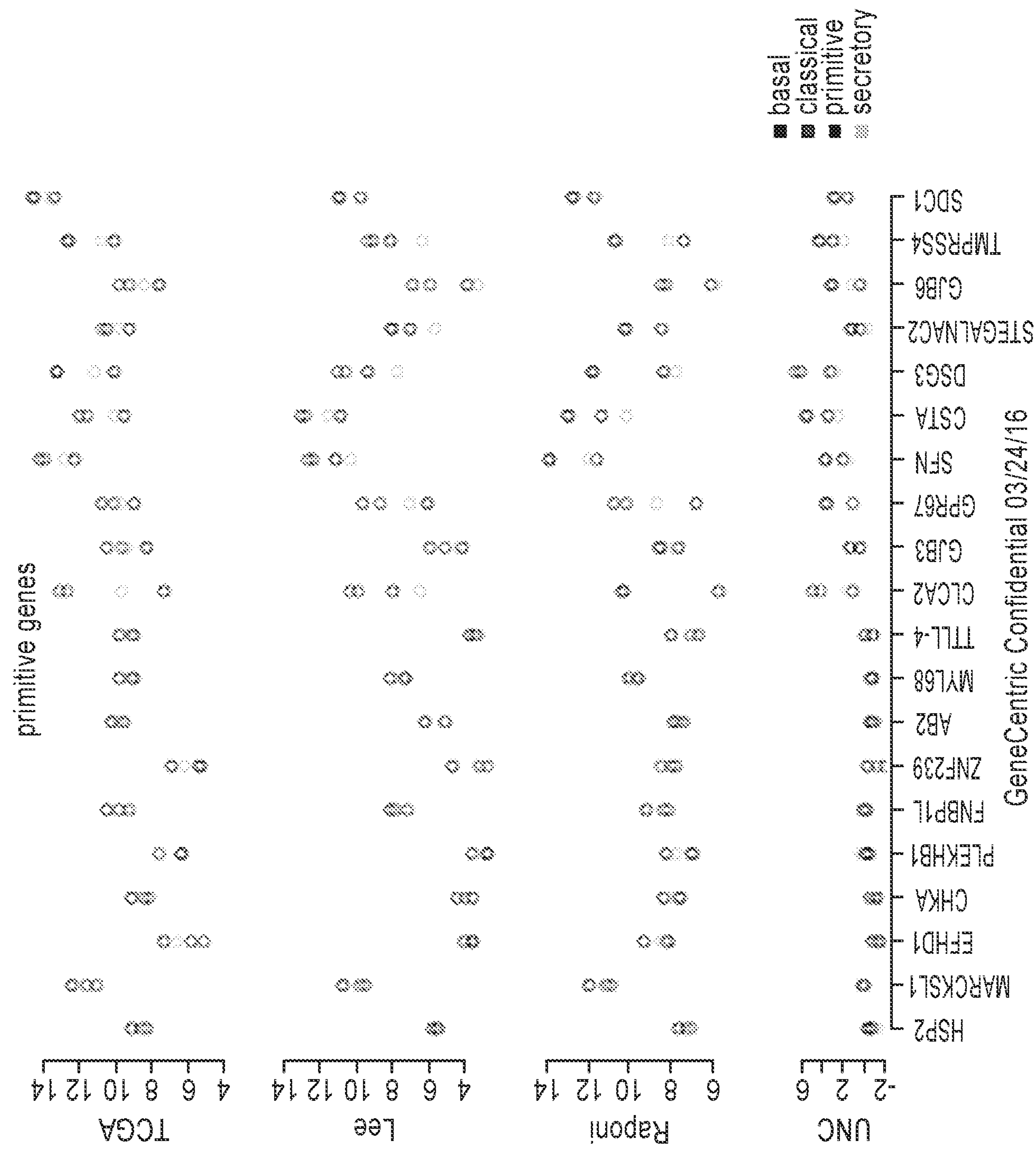


FIG. 14

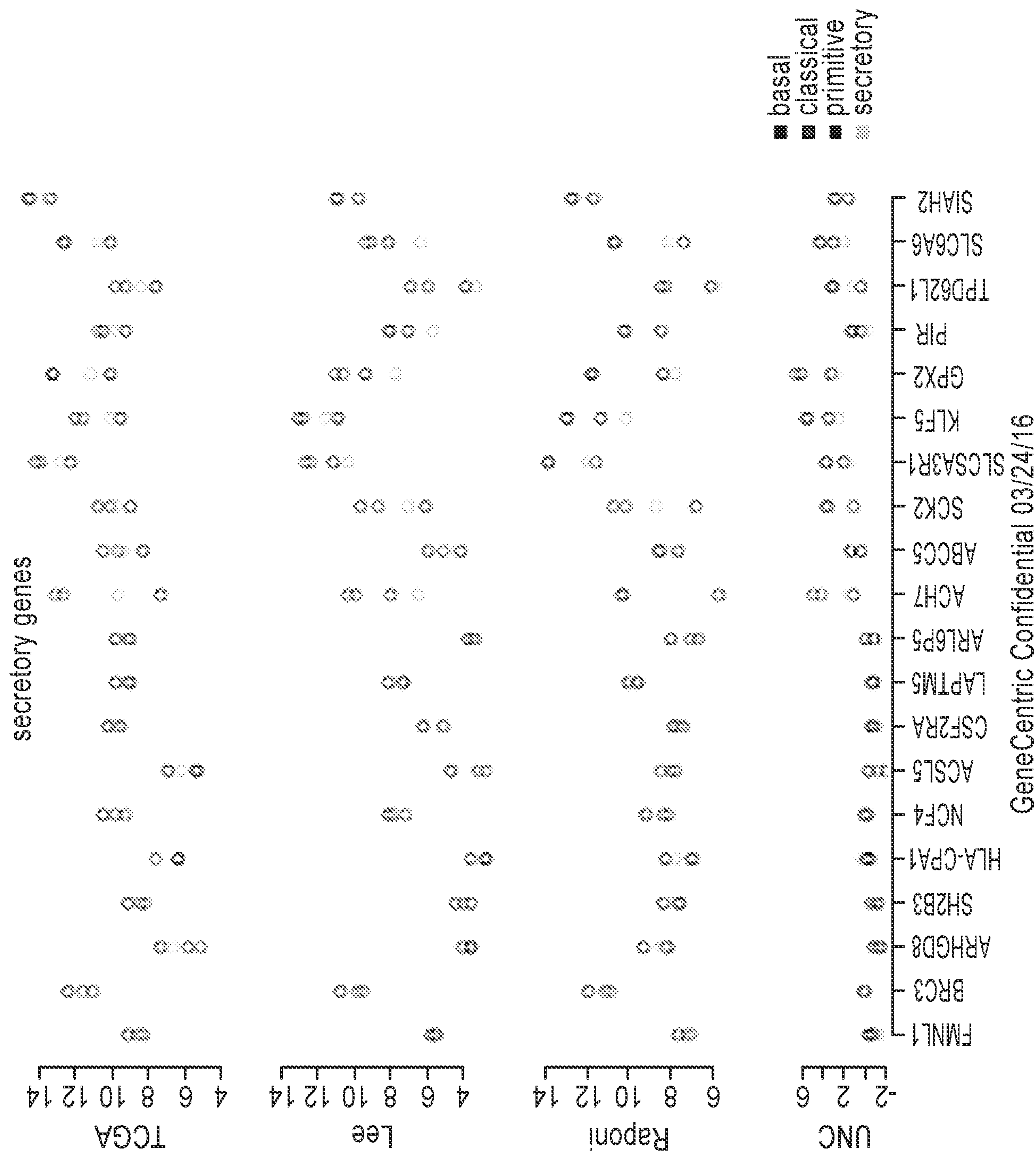


FIG. 15

18/36

## FIG. 16

Gold standard (rows) vs clanc80 (columns)

|                  |           | basal | classical | primitive | secretory |
|------------------|-----------|-------|-----------|-----------|-----------|
| Lee              | basal     | 13    | 0         | 0         | 0         |
|                  | classical | 2     | 29        | 0         | 0         |
|                  | primitive | 1     | 3         | 8         | 0         |
|                  | secretory | 5     | 1         | 0         | 13        |
| Raponi           | basal     | 35    | 0         | 0         | 0         |
|                  | classical | 0     | 42        | 0         | 0         |
|                  | primitive | 0     | 2         | 18        | 3         |
|                  | secretory | 4     | 0         | 3         | 23        |
| GeneCentric FFPE | basal     | 8     | 0         | 0         | 0         |
|                  | classical | 1     | 15        | 2         | 0         |
|                  | primitive | 0     | 0         | 5         | 0         |
|                  | secretory | 1     | 0         | 2         | 12        |
| TCGA             | basal     | 130   | 3         | 1         | 3         |
|                  | classical | 4     | 165       | 1         | 0         |
|                  | primitive | 2     | 6         | 56        | 9         |
|                  | secretory | 13    | 4         | 12        | 92        |
| UNC              | basal     | 12    | 0         | 0         | 0         |
|                  | classical | 1     | 20        | 0         | 0         |
|                  | primitive | 0     | 2         | 7         | 0         |
|                  | secretory | 3     | 0         | 0         | 11        |

Agreement

|       | Lee  | Raponi | GeneCentric<br>FFPE | TCGA | UNC  |
|-------|------|--------|---------------------|------|------|
| Agree | 0.84 | 0.91   | 0.87                | 0.88 | 0.89 |

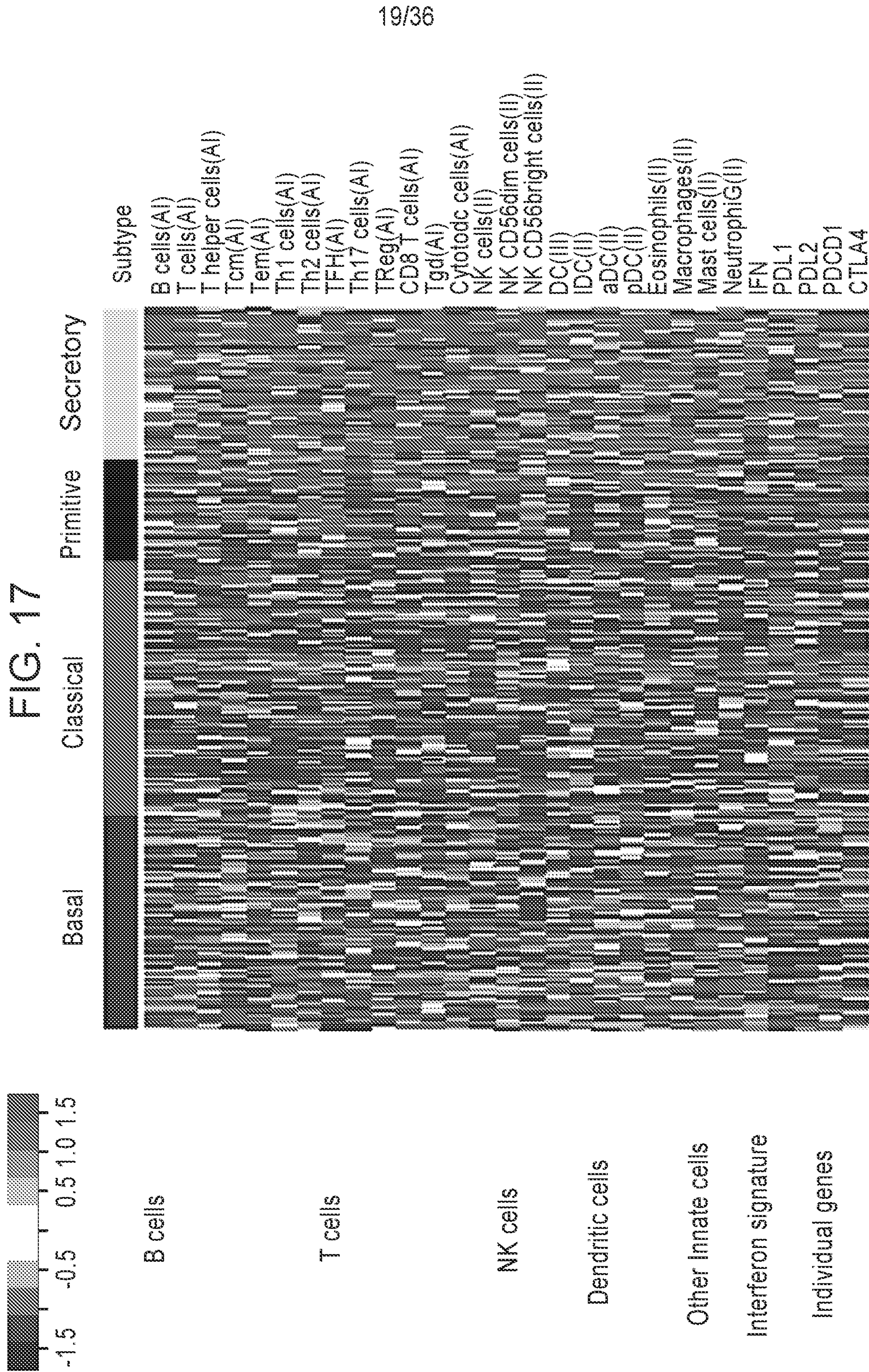


FIG. 18

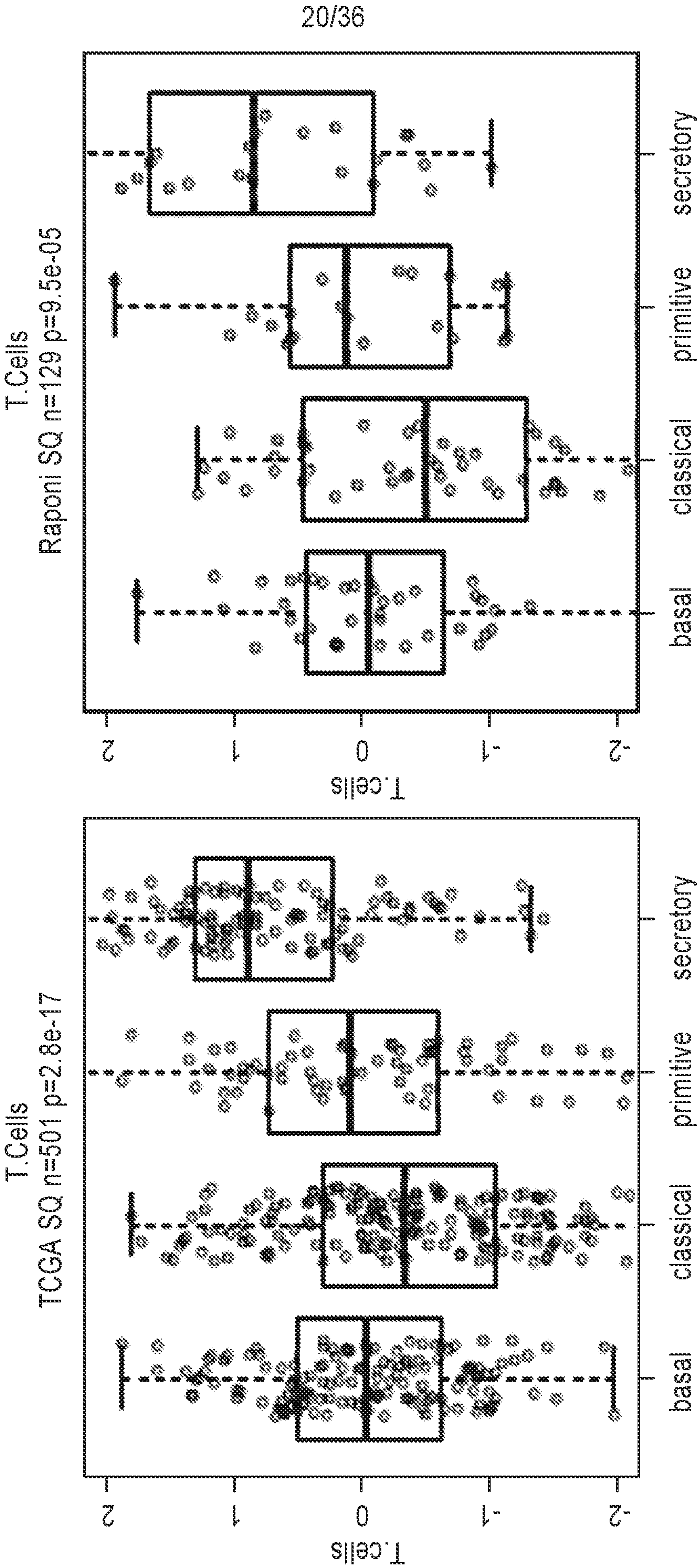
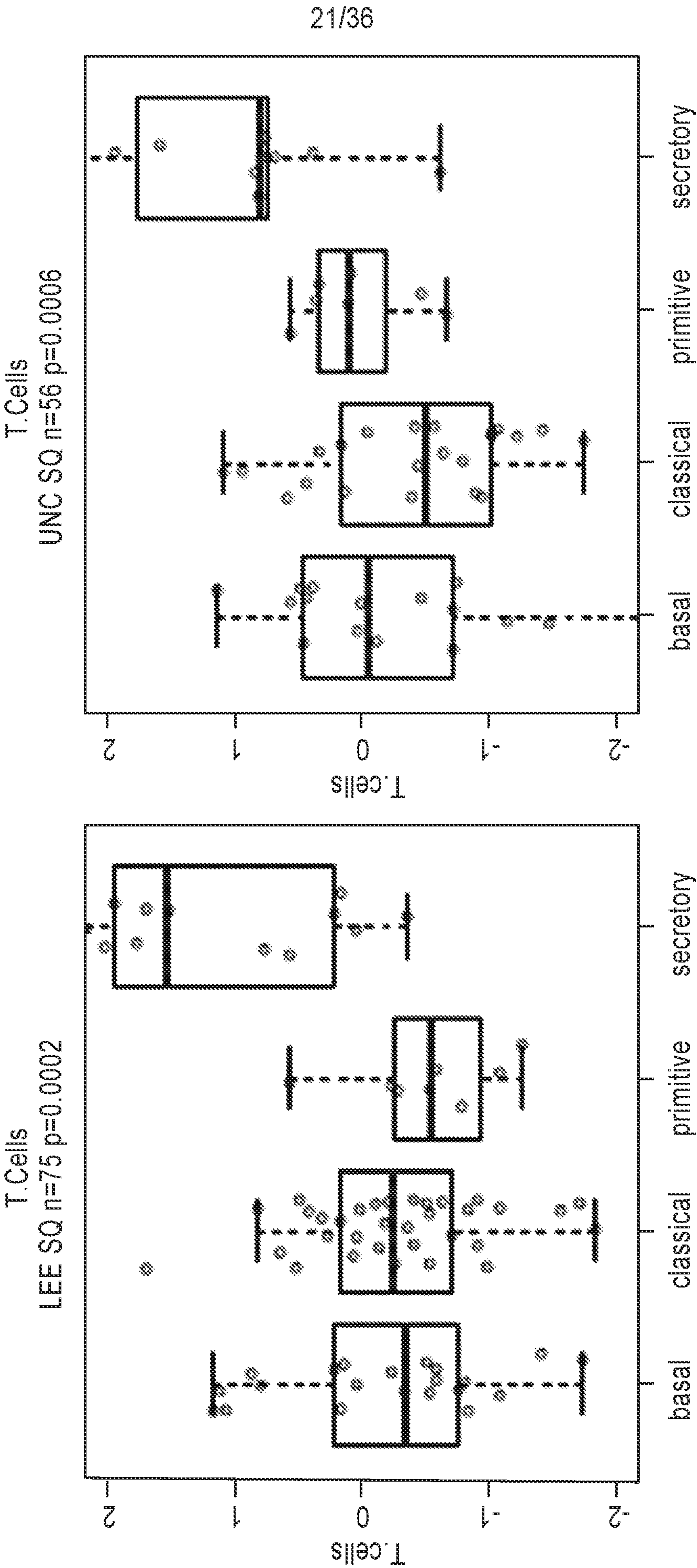
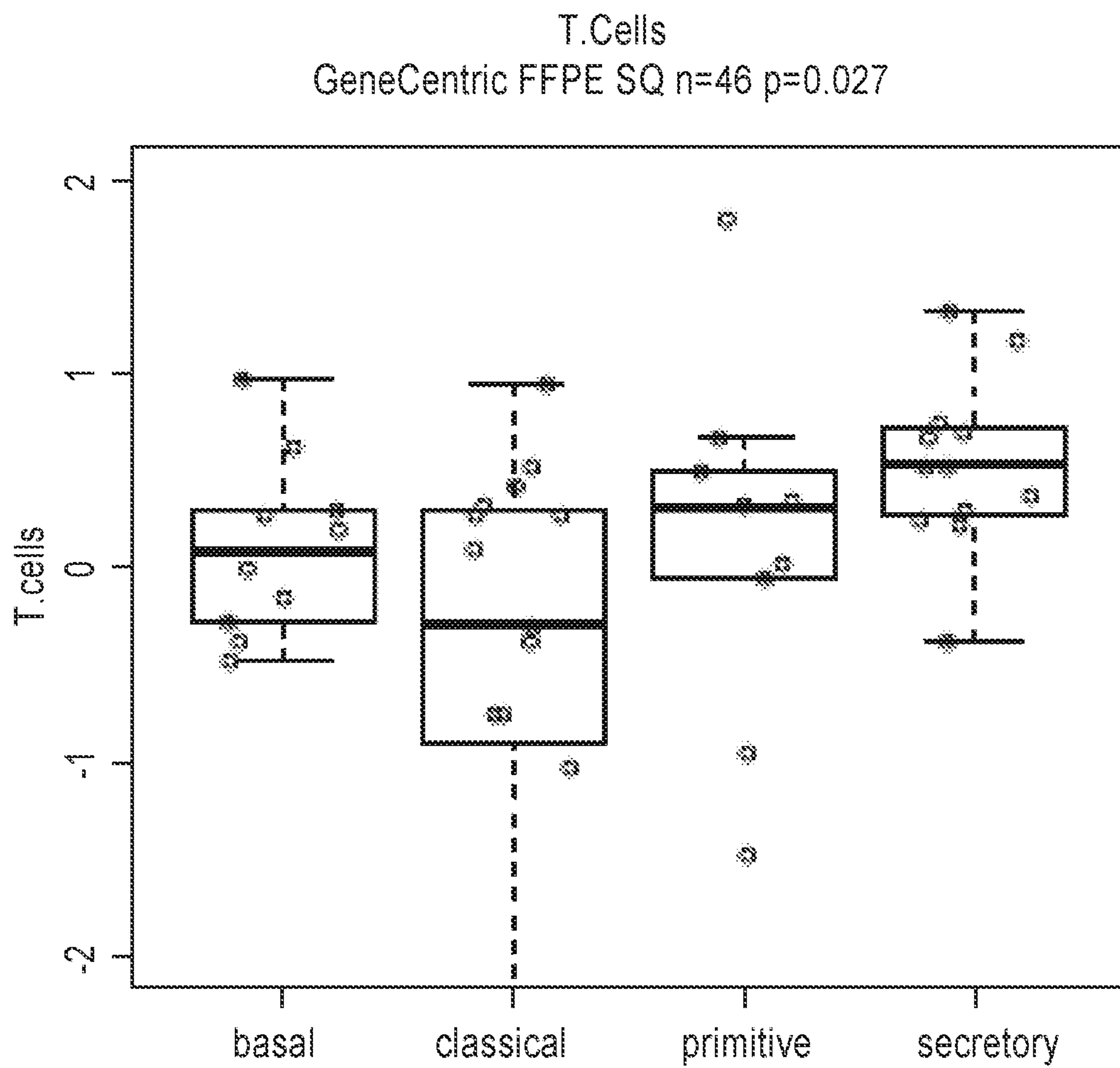


FIG. 18  
(continued)



22/36

FIG. 18  
(continued)

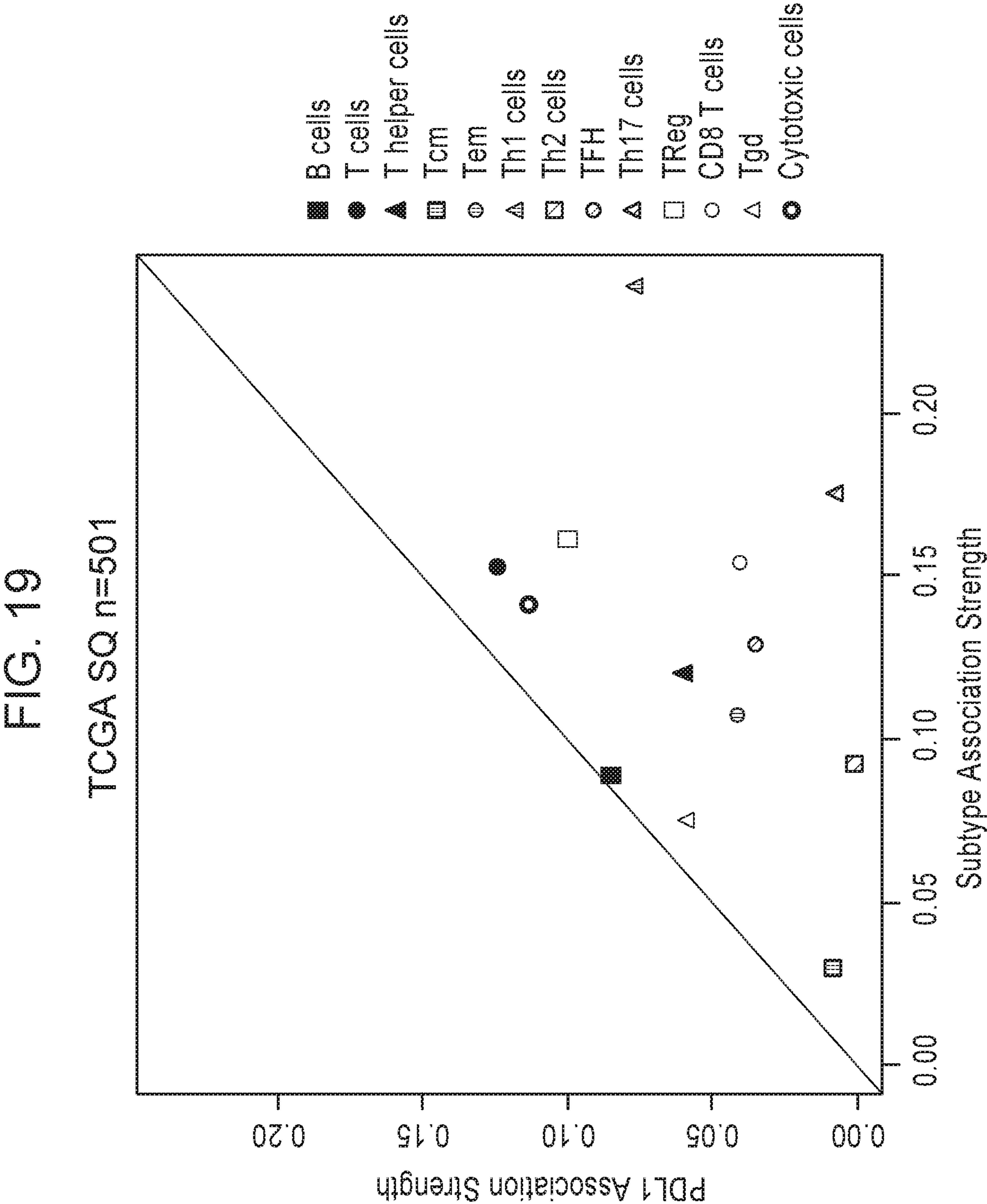
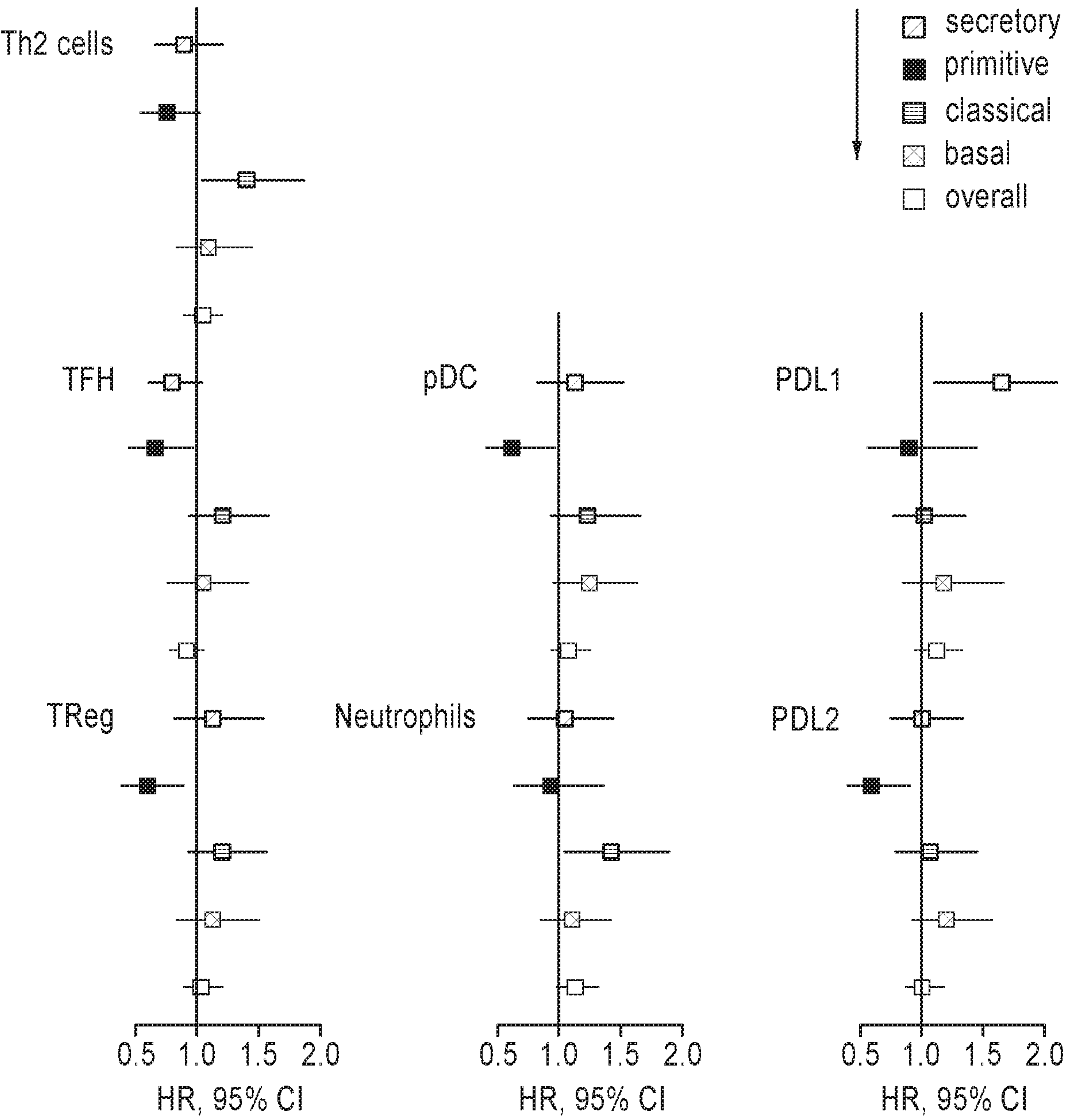
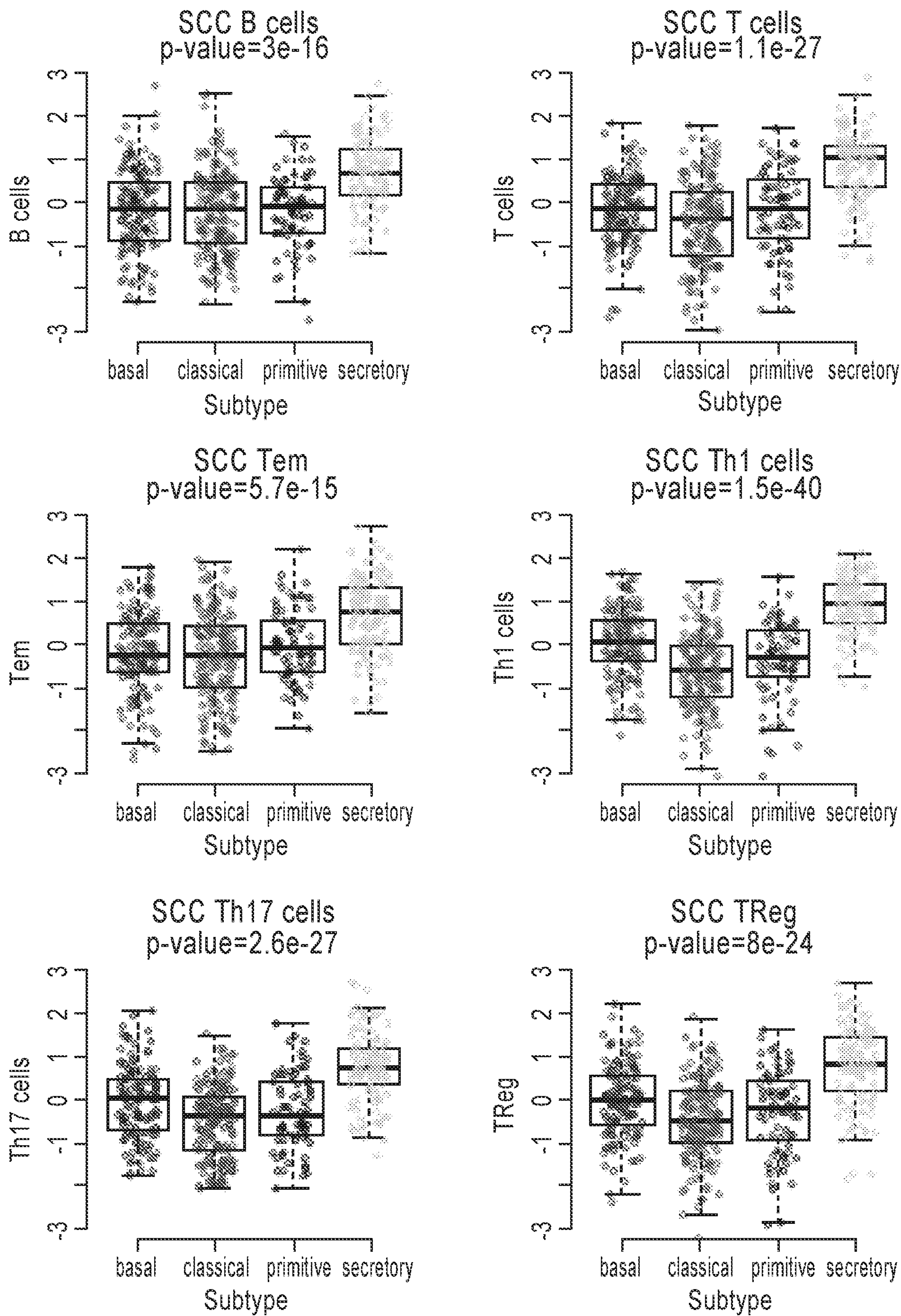


FIG. 20



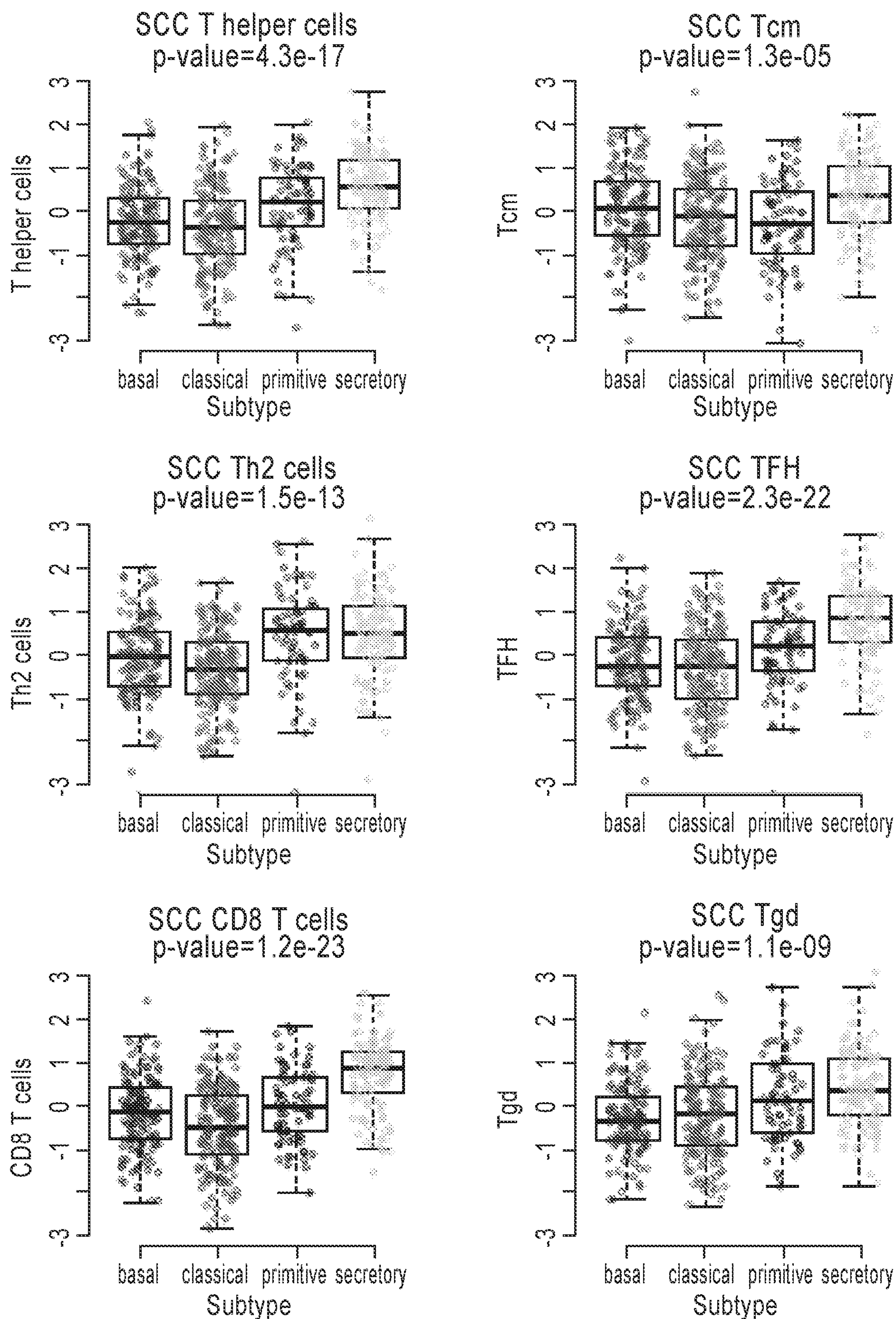
25/36

FIG. 21



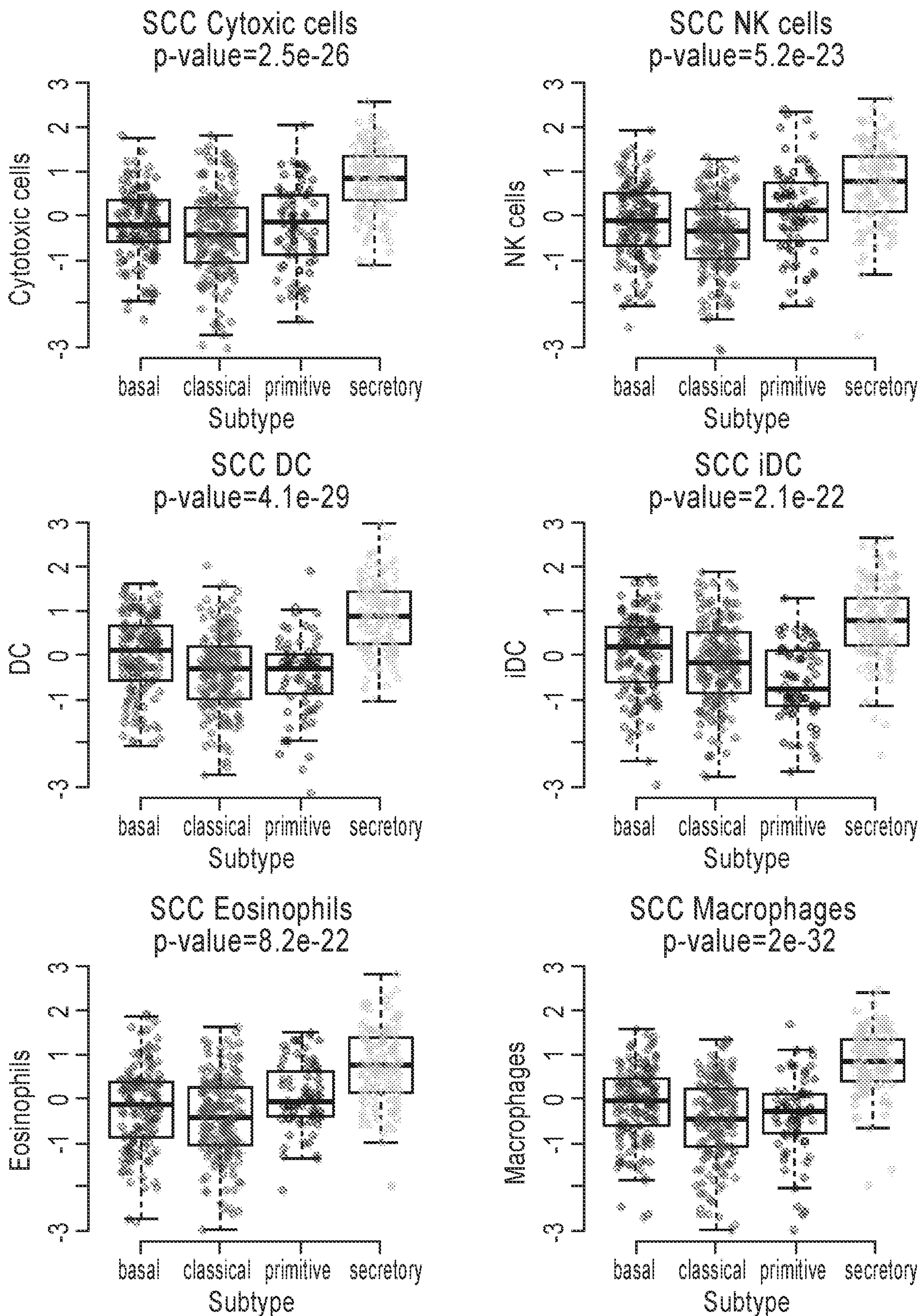
26/36

FIG. 21 (continued)



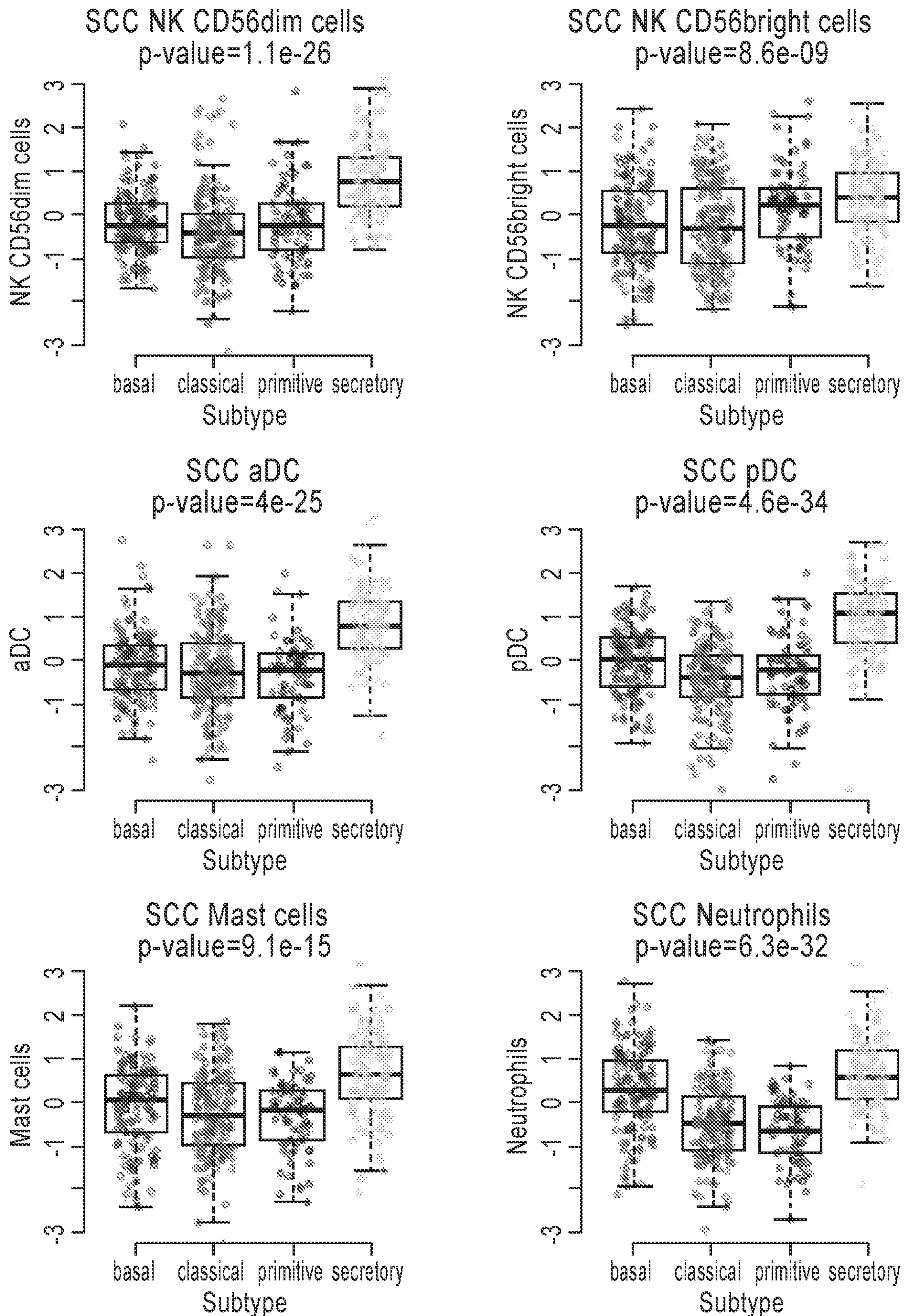
27/36

FIG. 21 (continued)



28/36

FIG. 21 (continued)



29/36

FIG. 21 (continued)

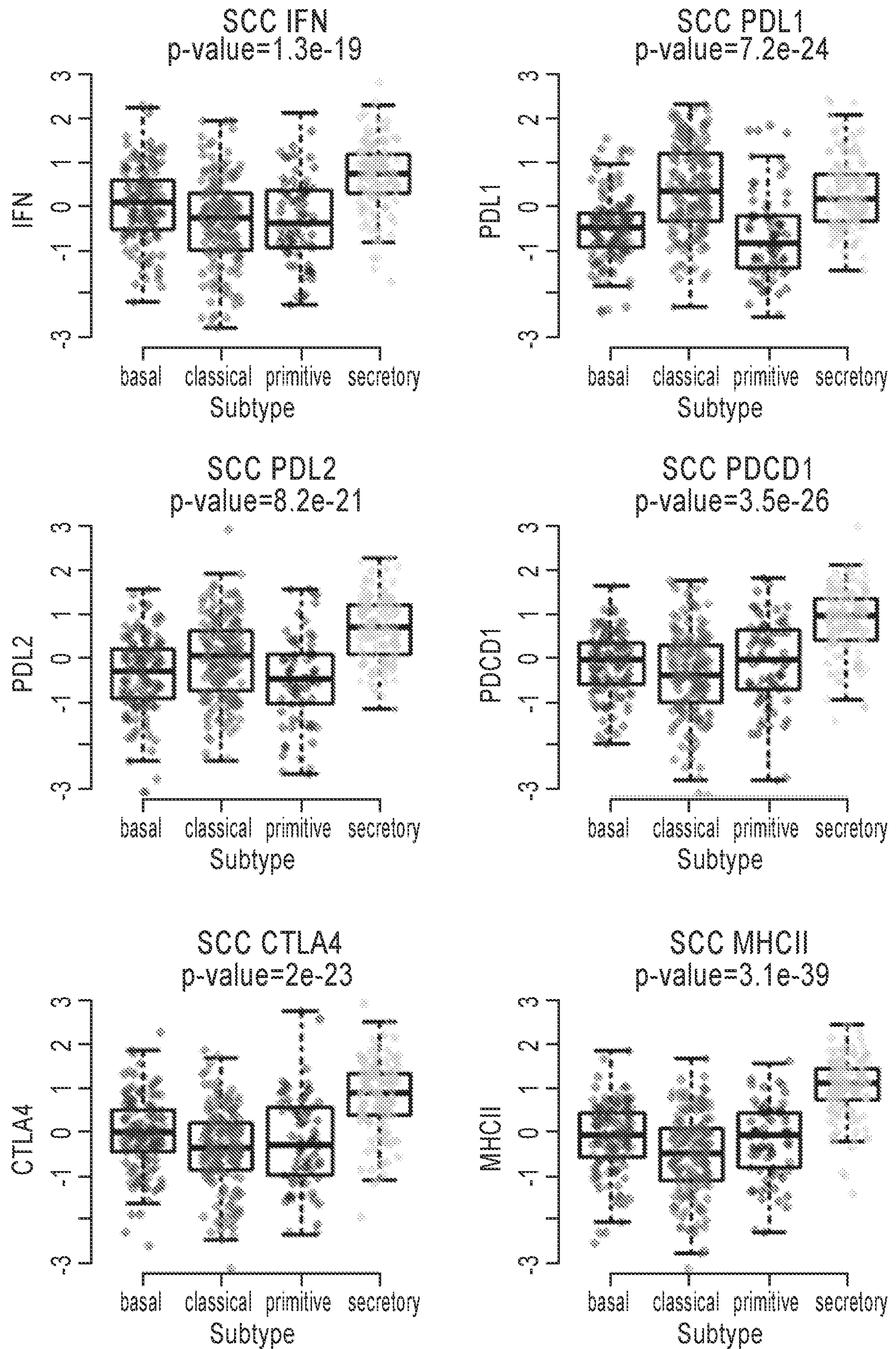


FIG. 22

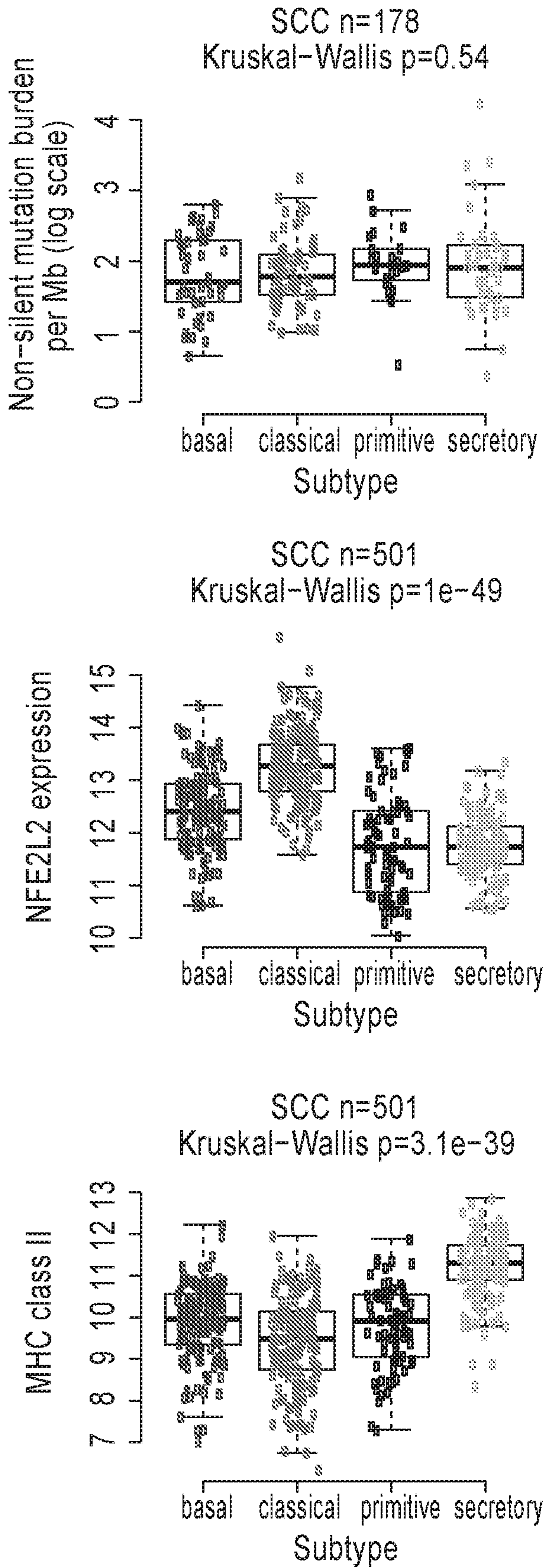


FIG. 23

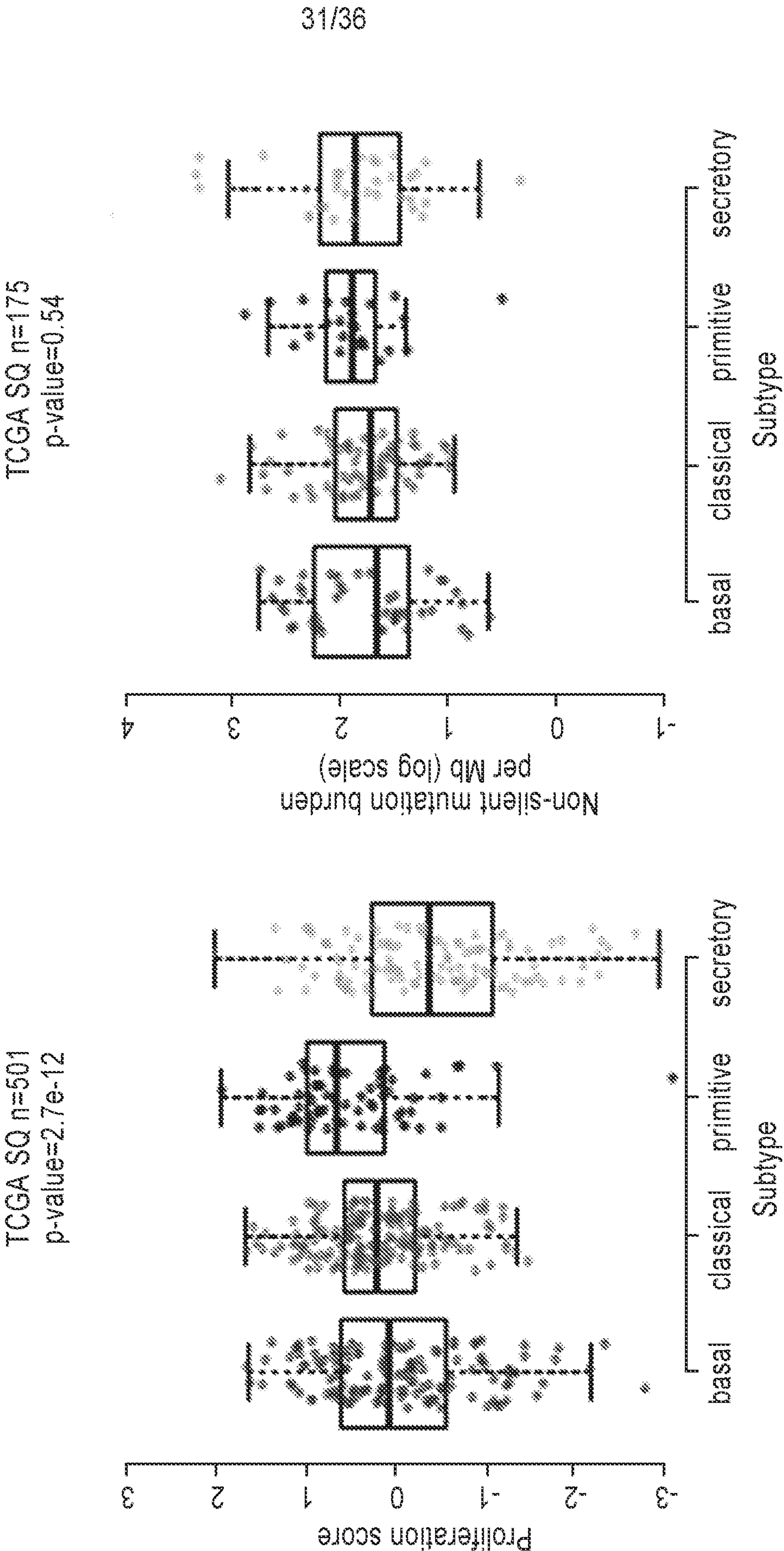
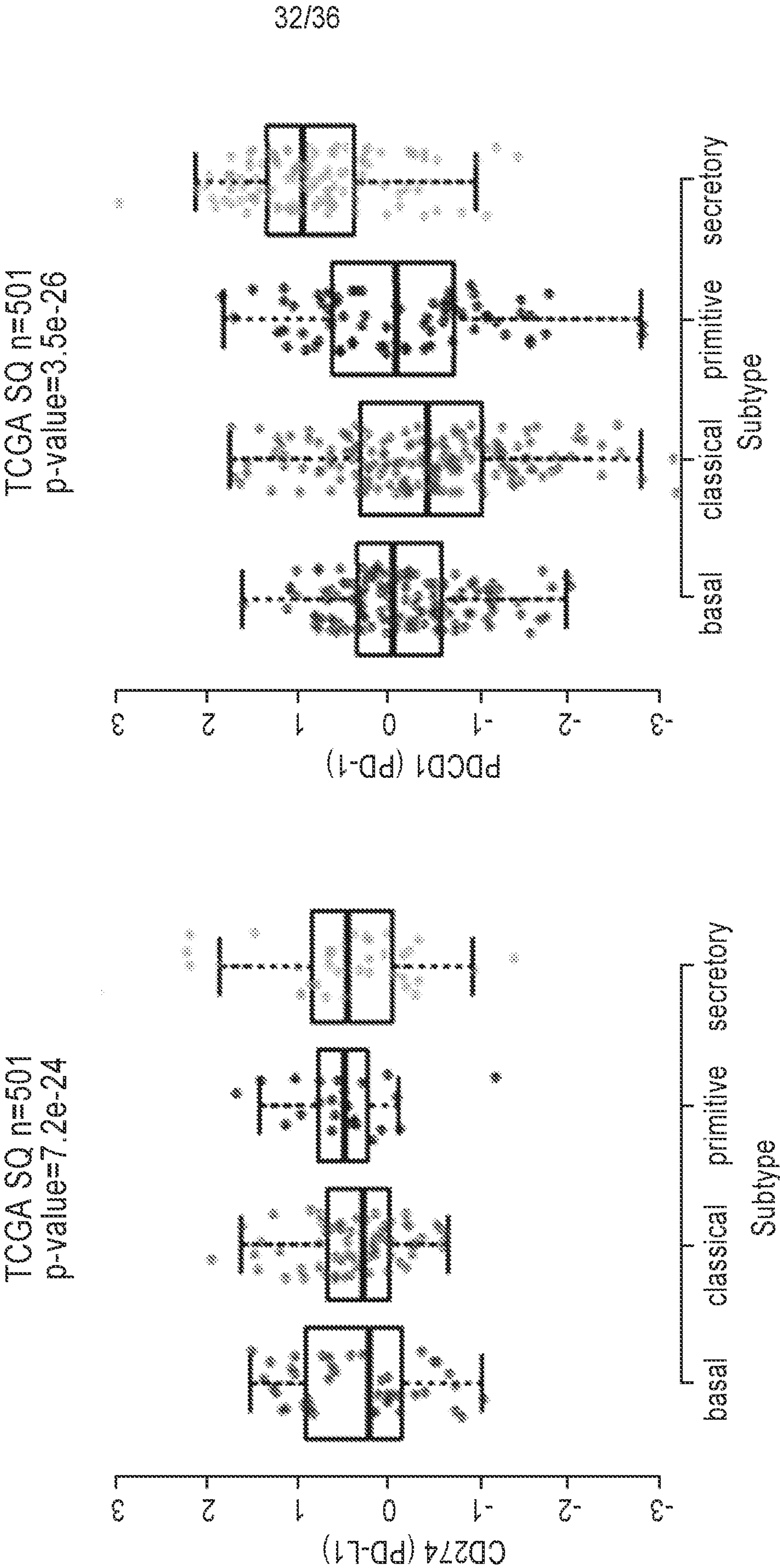


FIG. 23  
(continued)



33/36

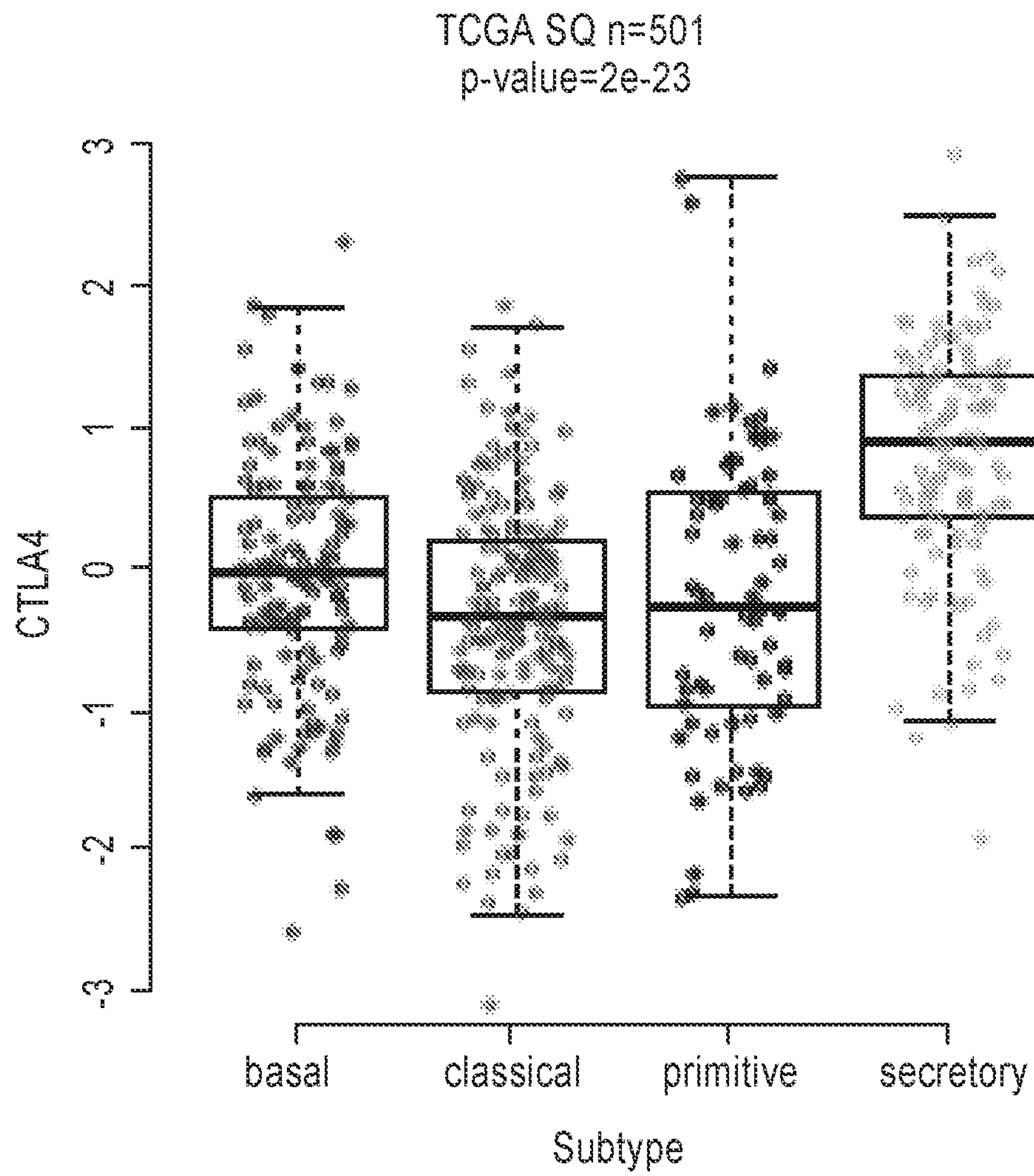
FIG. 23  
(continued)

FIG. 24

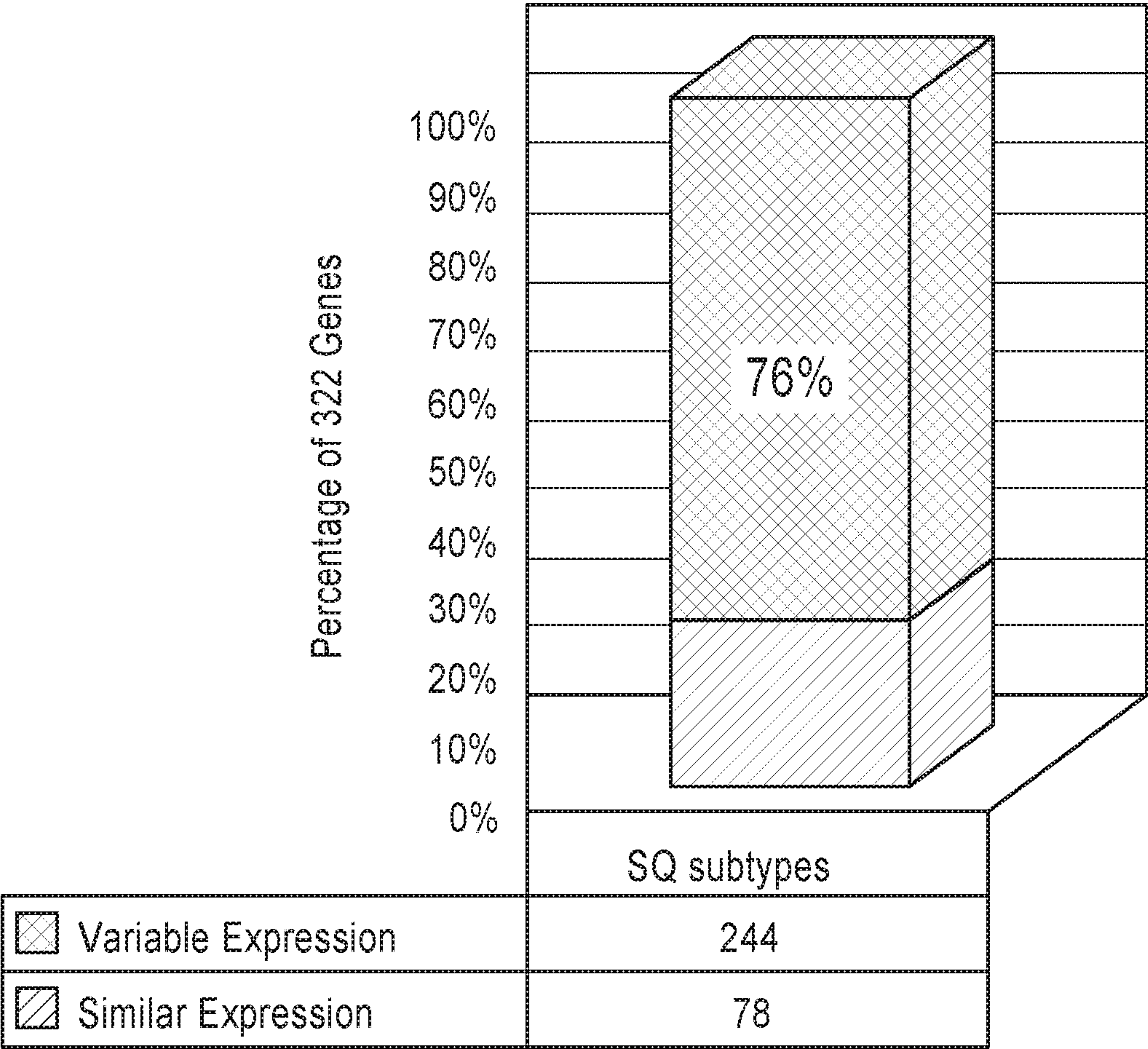
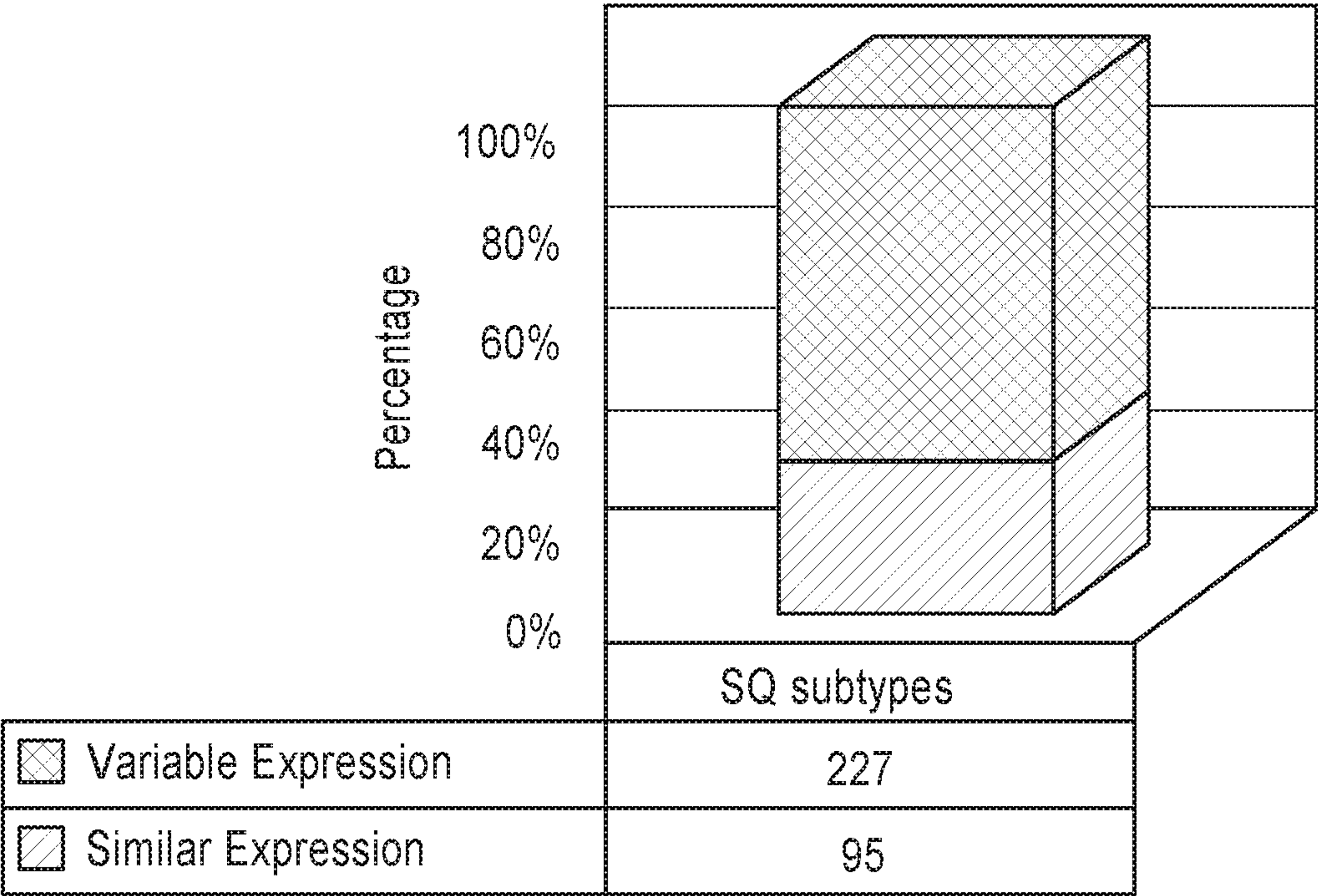


FIG. 25



36/36

FIG. 26

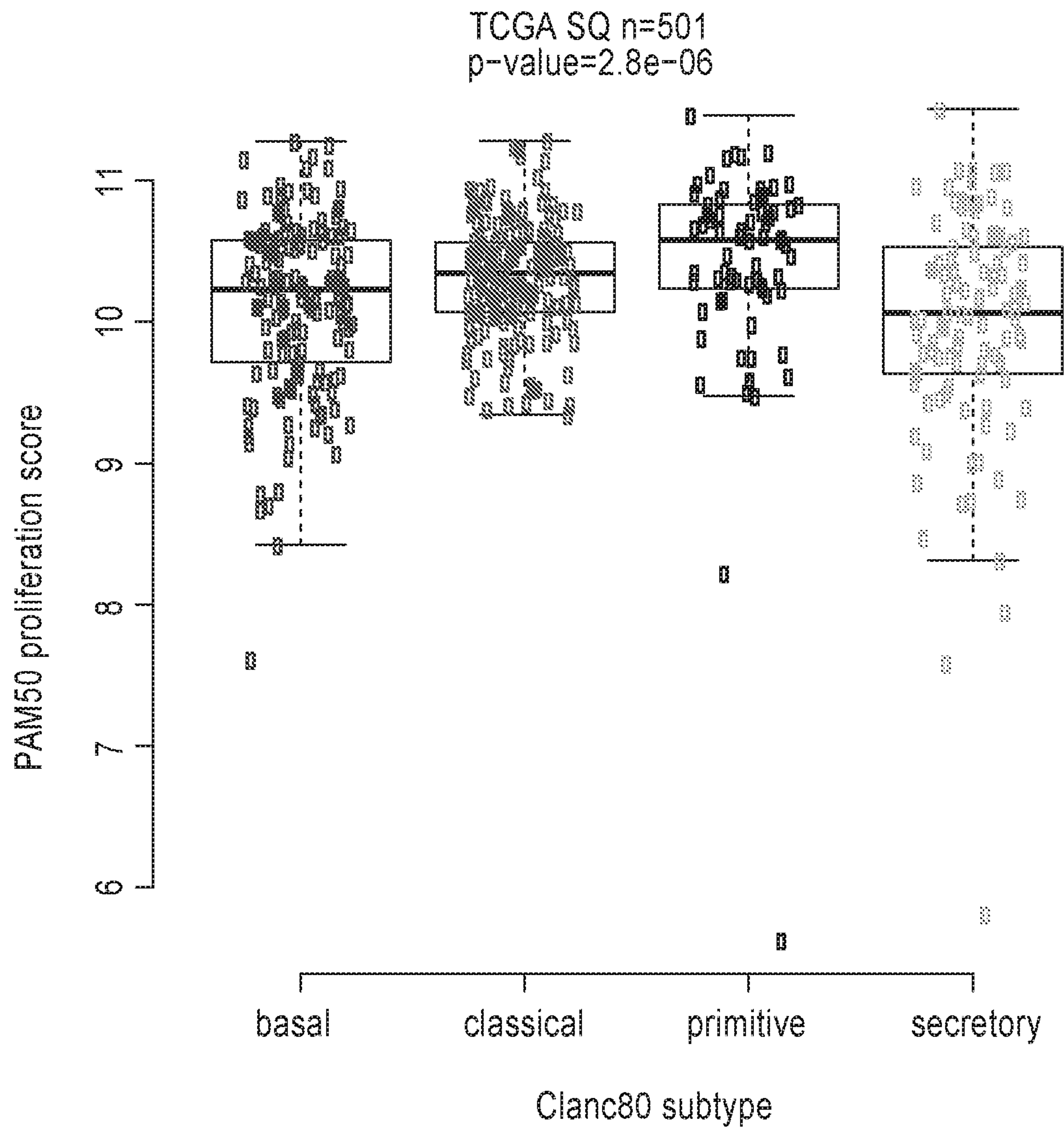


FIG. 1

## Lung Cancer Subtyping

