

TERTIARY LYMPHOID STRUCTURE COMPONENT MARKER COMBINATION AND SYSTEM FOR PREDICTING PROGNOSIS OF NASOPHARYNGEAL CANCER, AND USE

TECHNICAL FIELD

The present invention belongs to the field of biomedicine, particularly it relates to a marker combination, a system, and a use for tertiary lymphoid structure (TLS) components in predicting prognosis of nasopharyngeal carcinoma (NPC).

BACKGROUND

Nasopharyngeal carcinoma (NPC) is a unique type of head and neck tumor closely associated with Epstein-Barr virus (EBV) infection. It is highly prevalent in south China and is often referred to as "Canton tumor".

NPC originates in the nasopharynx with an insidious onset, typically being diagnosed at intermediate or advanced stages. Due to the inherent radiosensitivity of NPC tumor cells, radiotherapy serves as the primary treatment for NPC patients. Due to the technological advancements in radiotherapy and adjuvant chemotherapy, the survival rates of NPC patients have significantly improved, achieving a 5-year overall survival rate of over 85%. Nevertheless, more than 10% of NPC patients still develop recurrent or metastatic disease.

Patients at the same clinical stage may exhibit different prognosis after receiving identical treatments. This reflects the biological heterogeneity among tumors. Therefore, gaining a deeper understanding of the biological characteristics of NPC has emerged as a pivotal research focus.

In addition to highly heterogeneous tumor cells, NPC tumor tissues also contain infiltrating immune cells and other components at the lesion site. These cellular elements collectively constitute the tumor microenvironment, and differences in the components of the tumor microenvironment can influence the biological behavior of the tumor, potentially leading

to different prognosis.

Previous studies have reported that tertiary lymphoid structures (TLS)—regions where immune cells aggregate within tumors—can predict prognosis of the patients. However, the current research on TLS in NPC remains limited, and there is no systematic reporting on their role in predicting prognosis of NPC patients. A deeper analysis of TLS components in NPC could enhance our understanding of the disease's biological characteristics, help predict clinical prognosis, enable personalized treatment, and ultimately improve patient survival.

TLS are scattered within tumors. Conventional techniques have only been able to detect their presence at a morphological level, while their cellular composition and molecular profiles remain poorly understood. Accurately characterizing the cellular composition and molecular profiles of TLS still poses a significant challenge in current research.

SUMMARY

To address the above technical defects, an objective of the present invention is to provide a combination of tertiary lymphoid structure (TLS) component marker for predicting prognosis of nasopharyngeal carcinoma (NPC).

Another objective of the present invention is to provide a system for predicting prognosis of NPC.

A further objective of the present invention is to provide use of the aforementioned TLS component marker for predicting prognosis of NPC.

To achieve these objectives, the present invention provides a combination of the TLS component marker for predicting prognosis of NPC. The TLS component marker comprises one or more component markers as follows: germinal center B cell component marker, follicular helper CD4⁺ T cell component marker, CXCL13⁺ CD8⁺ T cell component marker, and CXCL13⁺ fibroblast component marker.

These four TLS component markers can be used individually or in any combination to predict prognosis of NPC.

Preferably, the germinal center B cell component marker comprises one or more characteristic genes or proteins as follows:

Regulator Of G Protein Signaling 13 (RGS13), MARCKS Like 1 (MARCKSL1), Nei Like DNA Glycosylase 1 (NEIL1), TCL1 Family AKT Coactivator A (TCL1A), High Mobility Group Nucleosome Binding Domain 1 (HMGN1), Inositol 1,4,5-Triphosphate Receptor Associated 2 (LRMP), Actin Gamma 1 (ACTG1), Serpin Family A Member 9 (SERPINA9), RP11-231C14.7, LIM Domain Containing Preferred Translocation Partner In Lipoma (LPP), CD79b Molecule (CD79B), ATP synthase subunit Lc (ATP5L), Small EDRK-Rich Factor 2 (SERF2), Succinyl-CoA:Glutarate-CoA Transferase (SUGCT), Lymphotoxin Beta (LTB), 5-Hydroxymethylcytosine Binding, ES Cell Specific (HMCES), Brain Abundant Membrane Attached Signal Protein 1 (BASP1), LIM Domain Only 2 (LMO2), Lipid Raft Linker 1 (RFTN1), Small Integral Membrane Protein 14 (SMIM14), BAF Chromatin Remodeling Complex Subunit BCL7A (BCL7A), Dishevelled Associated Activator Of Morphogenesis 1 (DAAM1), AC023590.1, BCL2 Interacting Killer (BIK), Actin Related Protein 2/3 Complex Subunit 2 (ARPC2), Poly(ADP-Ribose) Polymerase 1 (PARP1), GTP Cyclohydrolase I Feedback Regulator (GCHFR), uncoupling protein 2 (UCP2), Transcription Elongation Factor A1 (TCEA1), Methionine aminopeptidase 2 (METAP2), CD40, Glyceraldehyde-3-phosphate dehydrogenase (GAPDH), CD22, Breast Carcinoma Amplified Sequence 4 (BCAS4), Spectrin Repeat Containing Nuclear Envelope Protein 2 (SYNE2), LIM domain containing 2 (LIMD2), V-Set Pre-B Cell Surrogate Light Chain 3 (VPREB3), Coiled-coil domain containing 144A (CCDC144A), Transketolase (TKT), Solute Carrier Family 25 Member 5 (SLC25A5), Coronin-1A (CORO1A), High Mobility Group AT-Hook 1 (HMGA1), Interferon Regulatory Factor 8 (IRF8), STAG3 Cohesin Complex Component (STAG3), Ezrin (EZR), Cofilin 1 (CFL1), lymphocyte cytosolic protein 1 (LCP1), Actin Beta (ACTB), H3 histone, family 3A (H3F3A), CD81, Coiled-Coil Domain Containing 88A (CCDC88A), Glyoxylate And

Hydroxypyruvate Reductase (GRHPR), Sushi Domain Containing 3 (SUSD3), Vanin 2 (VNN2), Aminoacylase 3 (ACY3), Transmembrane Protein 123 (TMEM123), Kelch Like Family Member 6 (KLHL6), Phosphoribosyl Pyrophosphate Synthetase Associated Protein 2 (PRPSAP2), Peroxiredoxin 6 (PRDX6), Arachidonate 5-Lipoxygenase Activating Protein (ALOX5AP), GDP-Mannose 4,6-Dehydratase (GMDS), Dehydrogenase/Reductase 9 (DHRS9), AC079767.4, Profilin 1 (PFN1), WD Repeat Domain 66 (WDR66), Activation-Induced Cytidine Deaminase (AICDA), Golgi-Associated Gamma Adaptin Ear Containing ARF Binding Protein 2 (GGA2), BRICK1 Subunit of SCAR/WAVE Actin Nucleating Complex (BRK1), Actin Related Protein 2/3 Complex Subunit 1B (ARPC1B), CD53, Acidic Nuclear Phosphoprotein 32 Family Member B (ANP32B), Myocyte Enhancer Factor 2B (MEF2B), Actin Related Protein 2/3 Complex Subunit 5 (ARPC5), Differentially Expressed in FDCP 8 Homolog (DEF8), Actin Related Protein 3 (ACTR3), Ornithine Decarboxylase Antizyme 1 (OAZ1), Myosin IE (MYO1E), SWAP Switching B-Cell Complex 70 kDa Subunit (SWAP70), Actin Related Protein 2/3 Complex Subunit 4 (ARPC4), Pyruvate Kinase, Muscle Isozyme (PKM), SWI/SNF Related Matrix Associated Actin Dependent Regulator of Chromatin Subfamily B Member 1 (SMARCB1), HOP Homeobox (HOPX), Major Histocompatibility Complex, Class II, DM Beta (HLA-DMB), Lymphocyte-Specific Protein Tyrosine Kinase (LCK), Actin Related Protein 2/3 Complex Subunit 3 (ARPC3), Actin Related Protein 2/3 Complex Subunit 3 (SRSF9), 5-Hydroxytryptamine Receptor 3A (HTR3A), RAS Related Protein 2 (RRAS2), Methyl-CpG Binding Domain Protein 4 (MBD4), SEC14 Like Lipid Binding 1 (SEC14L1), Interleukin 4 Receptor (IL4R), DDB1 and CUL4 Associated Factor 12 (DCAF12), Lysophospholipase 1 (LYPLA1), Synaptosome Associated Protein 23 (SNAP23), Late Endosomal/Lysosomal Adaptor, MAPK and MTOR Activator 5 (LAMTOR5), Ubiquitin Conjugating Enzyme E2 J1 (UBE2J1), Coiled-Coil Domain Containing 69 (CCDC69), Major Histocompatibility Complex, Class II, DM Alpha (HLA-DMA), Regulator of G Protein Signaling 10 (RGS10), CDV3 Homolog (CDV3).

Preferably, the follicular helper CD4⁺ T cell component marker comprises one or more characteristic genes or proteins as follows:

C-X-C Motif Chemokine Ligand 13 (CXCL13), Integral Membrane Protein 2A (ITM2A), RP5-1028K7.2, Programmed Cell Death 1 (PD-1) (PDCD1), Thymocyte Selection Associated High Mobility Group Box Family Member 2 (TOX2), Chitinase 3 Like 2 (CHI3L2), Islet Cell Autoantigen 1 (ICA1), SH2 Domain Containing 1A (SH2D1A), Insulin Like Growth Factor Binding Protein 4 (IGFBP4), PAS Domain Containing Serine/Threonine Kinase (PASK), Prostate Cancer Associated Transcript 29 (PCAT29), Nuclear Receptor Subfamily 3 Group C Member 1 (NR3C1), RP11-455F5.5, Single-Pass Membrane and Coiled-Coil Domain Containing 4 (SMCO4), CD200, Neuromedin B (NMB), Thymocyte Selection Associated High Mobility Group Box (TOX), Protein Tyrosine Phosphatase Non-Receptor Type 13 (PTPN13), C-X-C Motif Chemokine Receptor 5 (CXCR5), Zinc Finger Protein 331 (ZNF331), Transcription Factor 7 (TCF7), T Cell Immunoreceptor With Ig And ITIM Domains (TIGIT), T Cell Immunoreceptor With Ig And ITIM Domains (FKBP5), Ring Finger Protein 19A (RNF19A), Linker for Activation of T Cells (LAT), MAGE Family Member H1 (MAGEH1), Chromosome 9 Open Reading Frame 16 (C9orf16), LIM Zinc Finger Domain Containing 1 (LIMS1), Fatty Acid Binding Protein 5 (FABP5), Protein Phosphatase 1 Catalytic Subunit Gamma (PPP1CC), Serglycin (SRGN), THADA Armadillo Repeat Containing (THADA), Carboxypeptidase M (CPM), Interleukin 6 Signal Transducer (IL6ST), Coronin 1B (CORO1B), ST8 Alpha-N-Acetyl-Neuraminide Alpha-2,8-Sialyltransferase 1 (ST8SIA1), Coactosin Like F-Actin Binding Protein 1 (COTL1), Coactosin Like F-Actin Binding Protein 1 (IKZF3), Solute Carrier Family 9 Member A9 (SLC9A9), Regulator of G Protein Signaling 2 (RGS2), Interleukin 6 Receptor (IL6R), MAF BZIP Transcription Factor (MAF), Chromosome 1 Open Reading Frame 228 (C1orf228), C-X-C Motif Chemokine Receptor 4 (CXCR4), FYN Proto-Oncogene (FYN), G Protein Subunit Gamma 4 (GNG4), Nuclear Factor of Activated T Cells 1 (NFATC1), Yippee Like 5 (YPEL5), TBC1 Domain Family Member 4 (TBC1D4), JunB Proto-Oncogene, AP-1 Transcription Factor Subunit (JUNB), T Cell Receptor Beta Constant 2 (TRBC2), Nucleobindin 2 (NUCB2), Dual Specificity Phosphatase 6 (DUSP6), Granzyme M (GZMM), Legumain (LGMN), Ribonuclease T2 (RNASET2), Tandem C2 Domains Nuclear Protein (TC2N), Teashirt Zinc Finger Homeobox 2 (TSHZ2), CD84, Interleukin 16 (IL16),

Sideroflexin 1 (SFXN1), T Cell Receptor Alpha Constant (TRAC), Signal Regulatory Protein Gamma (SIRPG), FYN Binding Protein (FYB), Lymphoid Restricted Membrane Protein (LRMP), CD40 Ligand (CD40LG), B and T Lymphocyte Associated (BTLA), Ras Homolog Family Member B (RHOB), POU Class 2 Homeobox Associating Factor 1 (POU2AF1), Nuclear Receptor Subfamily 4 Group A Member 2 (NR4A2), Nuclear Receptor Subfamily 4 Group A Member 2 (PPP2R5C), Abelson Helper Integration Site 1 (AHI1), SH3 Domain and Tetratricopeptide Repeats 1 (SH3TC1), Transcriptional Repressor GATA Binding 1 (TRPS1), DENN Domain Containing 2D (DENND2D), Ankyrin Repeat and SOCS Box Containing 2 (ASB2), KIAA1551, RAB27A, Stomatin (STOM), Protein Tyrosine Phosphatase Receptor Type C (PTPRC), Fatty Acid Amide Hydrolase 2 (FAAH2), Dual Specificity Phosphatase 2 (DUSP2), CD27, Siah E3 Ubiquitin Protein Ligase 2 (SIAH2), Ecotropic Viral Integration Site 2B (EVI2B), Family With Sequence Similarity 118 Member A (FAM118A), TSPY Like 2 (TSPYL2), SEPTIN6, TNF Superfamily Member 8 (TNFSF8), Zinc Finger and BTB Domain Containing 10 (ZBTB10), Sestrin 3 (SESN3), Pyrin and HIN Domain Family Member 1 (PYHIN1), TRAF Interacting Protein With Forkhead Associated Domain (TIFA), Protein Tyrosine Phosphatase Non-Receptor Type 2 (PTPN2), Signal Peptidase Complex Subunit 1 (SPCS1), Cornichon Family AMPA Receptor Auxiliary Protein 1 (CNIH1), Transmembrane Protein 2 (TMEM2), Solute Carrier Family 25 Member 46 (SLC25A46), H2A Histone Family Member Z (H2AFZ), Protein Tyrosine Phosphatase Non-Receptor Type 7 (PTPN7).

Preferably, the CXCL13⁺ CD8⁺ T cell component marker comprises one or more characteristic genes or proteins as follows:

C-X-C Motif Chemokine Ligand 13 (CXCL13), T Cell Receptor Beta Variable 6-5 (TRBV6-5), Thymocyte Selection Associated High Mobility Group Box (TOX), CD27, FXYD Domain Containing Ion Transport Regulator 6 (FXYD6), RP5-1028K7.2, Granzyme K (GZMK), CD74, Teashirt Zinc Finger Homeobox 2 (TSHZ2), Dual Specificity Phosphatase 4 (DUSP4), Insulin Like Growth Factor Binding Protein 4 (IGFBP4), SH2 Domain Containing 1A (SH2D1A), FBJ Murine Osteosarcoma Viral Oncogene Homolog (FOS), Branched Chain

Amino Acid Transaminase 1 (BCAT1), Adaptor Related Protein Complex 3 Subunit Sigma 1 (AP3S1), Chitinase 3 Like 2 (CHI3L2), Caveolin 1 (CAV1), Major Histocompatibility Complex, Class II, DP Beta 1 (HLA-DPB1), Ectodermal-Neural Cortex 1 (ENC1), GTP Binding Protein Overexpressed in Skeletal Muscle (GEM), Nuclear Receptor Subfamily 3 Group C Member 1 (NR3C1), Integral Membrane Protein 2A (ITM2A), Major Histocompatibility Complex, Class II, DP Alpha 1 (HLA-DPA1), Major Histocompatibility Complex, Class II, DQ Alpha 1 (HLA-DQA1), PAS Domain Containing Serine/Threonine Kinase (PASK), LanC Like Glutathione S-Transferase 2 (LANCL2), Granzyme M (GZMM), Fatty Acid Binding Protein 5 (FABP5), Cytotoxic and Regulatory T Cell Molecule (CRTAM), Phorbol-12-Myristate-13-Acetate-Induced Protein 1 (PMAIP1), Dual Specificity Phosphatase 1 (DUSP1), Integral Membrane Protein 2C (ITM2C), Vesicular Overexpressed in Cancer Prosurvival Protein 1 (VOPP1), Vasoactive Intestinal Peptide Receptor 2 (VIPR2), Transcription Factor 7 (TCF7), Cadherin Related Family Member 1 (CDHR1), Regulator of G Protein Signaling 2 (RGS2), CD82, Major Histocompatibility Complex, Class II, DR Alpha (HLA-DRA), Claudin Domain Containing 1 (CLDND1), Zinc Finger Protein 331 (ZNF331), Amyloid Beta Precursor Like Protein 2 (APLP2), Fc Receptor Like 3 (FCRL3), JunB Proto-Oncogene, AP-1 Transcription Factor Subunit (JUNB), Tubulin Alpha 1a (TUBA1A), SEPTIN6, TRAF Interacting Protein With Forkhead Associated Domain (TIFA), Chimerin 1 (CHN1), T Cell Receptor Associated Transmembrane Adaptor 1 (TRAT1), Androgen-Induced 1 (AIG1), Major Histocompatibility Complex, Class II, DR Beta 1 (HLA-DRB1), T Cell Receptor Alpha Variable 3 (TRAV3), Heat Shock Protein Family A (Hsp70) Member 1B (HSPA1B), Ras Homolog Family Member B (RHOB), Titin (TTN), C-X-C Motif Chemokine Receptor 5 (CXCR5), Dickkopf WNT Signaling Pathway Inhibitor 3 (DKK3), TIAM Rac1 Associated GEF 1 (TIAM1), Prostate Cancer Associated Transcript 29 (PCAT29), DnaJ Heat Shock Protein Family (Hsp40) Member B1 (DNAJB1), CD200, UDP-GlcNAc:BetaGal Beta-1,3-N-Acetylglucosaminyltransferase 2 (B3GNT2), Fatty Acid Amide Hydrolase 2 (FAAH2), Cadherin 1 (CDH1), Lymphocyte Antigen 9 (LY9), Eomesodermin (EOMES), CD8A, SEPTIN9, Phosphatidylethanolamine Binding Protein 1 (PEBP1), Signal Regulatory Protein

Gamma (SIRPG), Engulfment and Cell Motility 1 (ELMO1), Major Histocompatibility Complex, Class I, A (HLA-A), PDZ and LIM Domain 4 (PDLIM4), Sphingomyelin Phosphodiesterase Acid Like 3A (SMPDL3A), Linker for Activation of T Cells (LAT), Death Domain Containing 1 (DTHD1), C-X-C Motif Chemokine Receptor 3 (CXCR3), Major Histocompatibility Complex, Class II, DR Beta 5 (HLA-DRB5), Major Histocompatibility Complex, Class I, B (HLA-B), Paraoxonase 2 (PON2), Insulin Like Growth Factor Binding Protein 2 (IGFBP2), Stromal Antigen 3 (STAG3), Basic Leucine Zipper Transcription Factor, ATF-Like (BATF), Glutamate Receptor Ionotropic NMDA Associated Protein 1 (GRINA), Lysosomal Trafficking Regulator (LYST), Cold Inducible RNA Binding Protein (CIRBP), Interleukin 10 Receptor Subunit Alpha (IL10RA), Allograft Inflammatory Factor 1 (AIF1), Lipoprotein Lipase (LPL), Methyltransferase Like 8 (METTL8), Hypoxia Inducible Factor 1 Subunit Alpha (HIF1A), CD28, Phosphodiesterase 4D Interacting Protein (PDE4DIP), Semaphorin 4A (SEMA4A), ZNF1 Antisense RNA 1 (ZFAS1), L1 Cell Adhesion Molecule (L1CAM), MLLT3 Super Elongation Complex Subunit (MLLT3), Nerve Growth Factor Receptor Associated Protein 1 (NGFRAP1), TNF Receptor Superfamily Member 9 (TNFRSF9), Major Histocompatibility Complex, Class II, DQ Beta 1 (HLA-DQB1).

Preferably, the CXCL13+ fibroblast component marker comprises one or more the characteristic genes or proteins as follows:

Prostaglandin D2 Synthase (PTGDS), C-C Motif Chemokine Ligand 19 (CCL19), Matrix Metalloproteinase 1 (MMP1), Matrix Metalloproteinase 3 (MMP3), C-X-C Motif Chemokine Ligand 14 (CXCL14), C-X-C Motif Chemokine Ligand 1 (CXCL1), C-X-C Motif Chemokine Ligand 13 (CXCL13), Chitinase 3 Like 1 (CHI3L1), C-C Motif Chemokine Ligand 11 (CCL11), Complement Component 3 (C3), Retinoic Acid Receptor Responder 2 (RARRES2), C-C Motif Chemokine Ligand 21 (CCL21), Superoxide Dismutase 2 (SOD2), TNF Superfamily Member 13B (TNFSF13B), Complement C1s (C1S), Transmembrane Protein 176B (TMEM176B), Vascular Cell Adhesion Molecule 1 (VCAM1), ABI Family Member 3 Binding Protein (ABI3BP), Hydroxyprostaglandin Dehydrogenase (HPGD), Ceruloplasmin

(CP), Alcohol Dehydrogenase 1B (ADH1B), Dermopontin (DPT), CD24, C-C Motif Chemokine Ligand 2 (CCL2), C-X-C Motif Chemokine Ligand 6 (CXCL6), Transmembrane Protein 176A (TMEM176A), Metallothionein 2A (MT2A), Retinol Binding Protein 5 (RBP5), Lumican (LUM), Microfibril Associated Protein 4 (MFAP4), Decorin (DCN), Cytochrome P450 Family 1 Subfamily B Member 1 (CYP1B1), Fibroblast Growth Factor 7 (FGF7), Ferritin Heavy Chain 1 (FTH1), C-X-C Motif Chemokine Ligand 12 (CXCL12), Complement C1r (C1R), Iodothyronine Deiodinase 2 (DIO2), Cathepsin K (CTSK), Limbic System Associated Membrane Protein (LSAMP), Procollagen C-Endopeptidase Enhancer (PCOLCE), Follicular Dendritic Cell Secreted Protein (FDCSP), Rho Family GTPase 3 (RND3), Cathepsin S (CTSS), C-X-C Motif Chemokine Ligand 9 (CXCL9), Nuclear Receptor Subfamily 2 Group F Member 1 (NR2F1), Apolipoprotein E (APOE), Retinoic Acid Receptor Responder 1 (RARRES1), Lysyl Oxidase Like 1 (LOXL1), Flavin Containing Monooxygenase 2 (FMO2), Pleckstrin Homology Domain Containing H2 (PLEKHH2), Olfactomedin Like 3 (OLFML3), Coiled-Coil Domain Containing 80 (CCDC80), Twist Family BHLH Transcription Factor 2 (TWIST2), STEAP Family Member 1 (STEAP1), CD82, Serpin Family F Member 1 (SERPINF1), TNF Alpha Induced Protein 6 (TNFAIP6), Latent Transforming Growth Factor Beta Binding Protein 1 (LTBP1), Cellular Retinoic Acid Binding Protein 1 (CRABP1), Sclerostin Domain Containing 1 (SOSTDC1), Interleukin 32 (IL32), Proline/Arginine Rich End Leucine Rich Repeat Protein (PRELP), Indoleamine 2,3-Dioxygenase 1 (IDO1), TIMP Metallopeptidase Inhibitor 1 (TIMP1), Elastin Microfibril Interfacer 1 (EMILIN1), Podoplanin (PDPN), Selenoprotein M (SELM), Aldo-Keto Reductase Family 1 Member C1 (AKR1C1), Complement Component 7 (C7), AC090498.1, Complement Component 2 (C2), Insulin Like Growth Factor Binding Protein 5 (IGFBP5), Matrix Metallopeptidase 9 (MMP9), Tenascin C (TNC), G0/G1 Switch 2 (G0S2), Interferon Alpha Inducible Protein 27 Like 2 (IFI27L2), Prostaglandin E Synthase (PTGES), Platelet Derived Growth Factor Receptor Alpha (PDGFRA), Ubiquitin D (UBD), S100 Calcium Binding Protein A13 (S100A13), Clusterin (CLU), Isochorismatase Domain Containing 1 (ISOC1), Fibromodulin (FMOD), N-Myc Downstream Regulated 2 (NDRG2), Frizzled Related Protein (FRZB), Thymidine

Phosphorylase (TYMP), ADAM Like Decysin 1 (ADAMDEC1), Tetraspanin 8 (TSPAN8), Inter-Alpha-Trypsin Inhibitor Heavy Chain 5 (ITIH5), Galectin 3 Binding Protein (LGALS3BP), CXADR Like Membrane Protein (CLMP), Interleukin 34 (IL34), Complement Factor B (CFB), GTP Binding Protein Overexpressed in Skeletal Muscle (GEM), Inorganic Pyrophosphatase 1 (PPA1), Sequestosome 1 (SQSTM1), CD63, Transmembrane Protein 98 (TMEM98), CCAAT Enhancer Binding Protein Beta (CEBPB), Discoidin Domain Receptor Tyrosine Kinase 2 (DDR2).

In another aspect, the present invention also provides use of the aforementioned combination of TLS component marker in the preparation of a reagent for predicting prognosis of NPC.

In another aspect, the present invention also provides use of a product for quantifying the combination of TLS component marker in the preparation of the reagent for predicting prognosis of NPC.

Preferably, the product for quantifying the group of TLS component marker comprises a reagent for quantifying mRNA or protein levels of the TLS component marker. The product may also comprise other reagents, chips, or assays that reflect the expression level of mRNA or protein of the TLS component marker .

In another aspect, the present invention provides a system for predicting prognosis of NPC, comprising:

a TLS component marker quantification arrangement for determining an expression level of a combination of TLS component markers;

a data analysis arrangement for calculating based on the expression level of the combination of TLS component marker and determining prognosis of NPC;

a result output arrangement for outputting the calculated prognosis.

Preferably, the expression level may be determined by various methods, preferably the expression level processed by non-dimensional treatment, to further perform data analysis.

In some embodiments, the TLS component marker quantification arrangement comprises reagents, chips, detection tools, or kits for quantitatively detecting expression levels of mRNA or proteins of the marker.

In some embodiments, the arrangement for performing quantification of marker mRNA at the genetic level includes but is not limited to expression chips, mRNA sequencing tools, qPCR gene detection tools, and Nanostring technology gene detection tools. The tool for performing quantification of marker protein at the protein level includes but is not limited to enzyme-linked immunosorbent assay (ELISA), radioimmunoassay (RIA), immunohistochemistry, Western blot, electrophoresis, and liquid chromatography-tandem mass spectrometry (LC-MS/MS).

In some embodiments, the expression levels of each component can be calculated by performing single-sample gene set enrichment analysis (ssGSEA) on the expression levels of mRNA or protein for all characteristic genes of each component. The determined expression levels of mRNA or protein are normalized by different methods.

In some embodiments, if the expression level of a germinal center B cell component is below 10.77, it is classified as high risk; if the expression level of a germinal center B cell component is above 10.77, it is classified as low risk.

In some embodiments, if the expression level of a follicular helper CD4⁺ T-cell component is below 9.09, it is classified as high risk; if the expression level of a follicular helper CD4⁺ T cell component is above 9.09, it is classified as low risk.

In some embodiments, if the expression level of a CXCL13⁺ CD8⁺ T-cell component is below 8.41, it is classified as high risk; if the expression level of a CXCL13⁺ CD8⁺ T cell component is above 8.41, it is classified as low risk.

In some embodiments, if the expression level of a CXCL13+ fibroblast component is below 8.58, it is classified as high risk; if the expression level of a CXCL13+ fibroblast component is above 8.58, it is classified as low risk.

In some embodiments, other well-known methods may be used to determine specific risk thresholds.

By calculating the expression levels of the characteristic genes or proteins of these four immune cell components, the present invention enables predicting the expression levels of different TLS components in NPC and assess clinical prognosis, facilitating personalized treatment and ultimately improving patient survival rates.

BRIEF DESCRIPTION

Figure 1 depicts the normalized mean expression of classical TLS marker genes (columns) corresponding to different cell types (rows), wherein the color gradient ranging from blue (low) to red (high) represents the normalized mean expression levels.

Figure 2 shows the characteristic scores of germinal center B cell components, follicular helper CD4+ T cell components, CXCL13+ CD8+ T cell components, and CXCL13+ fibroblast components (rows) in 150 NPC samples (columns), wherein the color gradient ranging from blue to red indicates scaled expression levels from low to high, the colored bars at the top of the heatmap represent overall survival and progression-free survival information of the samples, and all samples are classified into two groups based on low- or high-level TLS.

Figure 3 shows the stratification of NPC patients into high-TLS and low-TLS groups based on TLS levels, wherein Kaplan-Meier survival curve analysis revealed distinct progression-free survival (PFS, left panel) and overall survival (OS, right panel) outcomes between these groups.

Figure 4 shows patient stratification into high-risk and low-risk groups according to the abundance of germinal center B cell components, follicular helper CD4⁺ T cell components, CXCL13⁺ CD8⁺ T cell components, and CXCL13⁺ fibroblast components, wherein Kaplan-Meier analysis demonstrated significant differences in PFS (left panel) and OS (right panel) between the stratified cohorts.

DETAILED DESCRIPTION

To provide a clearer understanding of the objectives, features, and advantages of the present invention, the following detailed description is provided with reference to the accompanying drawings and specific embodiments. It should be understood by those one skilled in the art that the described embodiments are intended to aid in comprehending the present invention and should not be construed as specific limitations thereof. Unless otherwise stated, the sequencing and analytical methods used in the following embodiments are standard protocols.

The invention is further illustrated below through specific experiments. It is understood that the following descriptions are provided solely to explain the present invention and do not limit the protection scope of the present invention.

By using single-cell mRNA sequencing technology and analyzing classical TLS marker genes (AICDA, RGS13, BCL6, CXCR5, CXCL13, CCL19, CCL21, PDCD1), it was discovered that TLSs in nasopharyngeal carcinoma (NPC) comprise four cell components: germinal center B cells, follicular helper CD4⁺ T cells, CXCL13⁺ CD8⁺ T cells, and CXCL13⁺ fibroblasts. Systematic bioinformatics analysis identified characteristic genes corresponding to each cell component. By measuring expression levels of mRNA corresponding to these characteristic genes in NPC tumor tissues and performing single-sample gene set enrichment analysis (ssGSEA), the relative abundance of each component in NPC tissues could be predicted. Furthermore, survival analysis revealed that the expression levels of these four cell components are predictive of clinical prognosis in NPC patients.

Cell components of TLS and their corresponding characteristic genes:

By using single-cell mRNA sequencing technology, the mRNA expression profiles of 349,038 cells from 67 NPC patients (obtained from Sun Yat-sen University Cancer Center) were analyzed. Clustering analysis was performed to reveal nine major cell subpopulations. Further sub-clustering of these subpopulations, combined with expression of TLS characteristic genes, identified four component markers originating from TLS: germinal center B cells, follicular helper CD4⁺ T cells, CXCL13⁺ CD8⁺ T cells, and CXCL13⁺ fibroblasts.

As shown in Figure 1, the normalized mean expression of classical TLS characteristic genes (columns) corresponding to different cell types (rows) is visualized. The color gradient ranging from blue to red represents low to high normalized mean expression levels. Figure 1 demonstrates that the characteristic genes of four cell components—germinal center B cells (RGS13⁺ B cells), follicular helper CD4⁺ T cells, CXCL13⁺ CD8⁺ T cells, and CXCL13⁺ fibroblasts—exhibit the highest expression levels.

Subsequently, differential gene analysis using the Wilcoxon rank-sum test was performed on these four cell components. The top 100 most significantly differentially expressed genes (ranked by fold change) were selected as characteristic genes for each component, establishing predictive markers for these cell components.

Characteristic genes for each component are shown in Table 1 as follows.

Table 1. Characteristic Genes of the Four Cell types

Cell type	Characteristic Genes
germinal center B cell	RGS13, MARCKSL1, NEIL1, TCL1A, HMGN1, LRMP, ACTG1, SERPINA9, RP11-231C14.7, LPP, CD79B, ATP5L, SERF2, SUGCT, LTB, HMCES, BASP1, LMO2, RFTN1, SMIM14, BCL7A, DAAM1, AC023590.1, BIK,

	ARPC2, PARP1, GCHFR, UCP2, TCEA1, METAP2, CD40, GAPDH, CD22, BCAS4, SYNE2, LIMD2, VPREB3, CCDC144A, TKT, SLC25A5, CORO1A, HMGA1, IRF8, STAG3, EZR, CFL1, LCP1, ACTB, H3F3A, CD81, CCDC88A, GRHPR, SUSP3, VNN2, ACY3, TMEM123, KLHL6, PRPSAP2, PRDX6, ALOX5AP, GMDS, DHRS9, AC079767.4, PFN1, WDR66, AICDA, GGA2, BRK1, ARPC1B, CD53, ANP32B, MEF2B, ARPC5, DEF8, ACTR3, OAZ1, MYO1E, SWAP70, ARPC4, PKM, SMARCB1, HOPX, HLA-DMB, LCK, ARPC3, SRSF9, HTR3A, RRAS2, MBD4, SEC14L1, IL4R, DCAF12, LYPLA1, SNAP23, LAMTOR5, UBE2J1, CCDC69, HLA-DMA, RGS10, CDV3
follicular helper CD4+ T cell	CXCL13, ITM2A, RP5-1028K7.2, PDCD1, TOX2, CHI3L2, ICA1, SH2D1A, IGFBP4, PASK, PCAT29, NR3C1, RP11-455F5.5, SMCO4, CD200, NMB, TOX, PTPN13, CXCR5, ZNF331, TCF7, TIGIT, FKBP5, RNF19A, LAT, MAGEH1, C9orf16, LIMS1, FABP5, PPP1CC, SRGN, THADA, CPM, IL6ST, CORO1B, ST8SIA1, COTL1, IKZF3, SLC9A9, RGS2, IL6R, MAF, C1orf228, CXCR4, FYN, GNG4, NFATC1, YPEL5, TBC1D4, JUNB, TRBC2, NUCB2, DUSP6, GZMM, LGMN, RNASET2, TC2N, TSHZ2, CD84, IL16, SFXN1, TRAC, SIRPG, FYB, LRMP, CD40LG, BTLA, RHOB, POU2AF1, NR4A2, PPP2R5C, AHI1, SH3TC1, TRPS1, DENND2D, ASB2, KIAA1551, RAB27A, STOM, PTPRC, FAAH2, DUSP2, CD27, SIAH2, EVI2B, FAM118A, TSPYL2, SEPTIN6, TNFSF8, ZBTB10, SESN3, PYHIN1, TIFA, PTPN2, SPCS1, CNIH1, TMEM2, SLC25A46, H2AFZ, PTPN7
CXCL13+ CD8+ T cell	CXCL13, TRBV6-5, TOX, CD27, FXVD6, RP5-1028K7.2, GZMK, CD74, TSHZ2, DUSP4, IGFBP4, SH2D1A, FOS, BCAT1, AP3S1, CHI3L2, CAV1, HLA-DPB1, ENC1, GEM, NR3C1, ITM2A, HLA-DPA1, HLA-DQA1, PASK, LANCL2, GZMM, FABP5, CRTAM, PMAIP1, DUSP1, ITM2C, VOPP1, VIPR2, TCF7, CDHR1, RGS2, CD82, HLA-DRA, CLDN1, ZNF331, APLP2, FCRL3, JUNB, TUBA1A, SEPTIN6, TIFA, CHN1, TRAT1, AIG1, HLA-DRB1, TRAV3, HSPA1B, RHOB, TTN, CXCR5, DKK3, TIAM1, PCAT29, DNAJB1, CD200, B3GNT2, FAAH2, CDH1, LY9, EOMES, CD8A, SEPTIN9, PEBP1, SIRPG, ELMO1, HLA-A, PDLIM4, SMPDL3A, LAT, DTHD1, CXCR3, HLA-DRB5, HLA-B, PON2, IGFBP2, STAG3, BATF, GRINA, LYST, CIRBP, IL10RA, AIF1, LPL, METTL8, HIF1A, CD28, PDE4DIP, SEMA4A, ZFAS1, L1CAM, MLLT3, NGFRAP1, TNFRSF9, HLA-DQB1
CXCL13+ fibroblast	PTGDS, CCL19, MMP1, MMP3, CXCL14, CXCL1, CXCL13, CHI3L1, CCL11, C3, RARRES2, CCL21,

	SOD2, TNFSF13B, C1S, TMEM176B, VCAM1, ABI3BP, HPGD, CP, ADH1B, DPT, CD24, CCL2, CXCL6, TMEM176A, MT2A, RBP5, LUM, MFAP4, DCN, CYP1B1, FGF7, FTH1, CXCL12, C1R, DIO2, CTSK, LSAMP, PCOLCE, FDCSP, RND3, CTSS, CXCL9, NR2F1, APOE, RARRES1, LOXL1, FMO2, PLEKHH2, OLFML3, CCDC80, TWIST2, STEAP1, CD82, SERPINF1, TNFAIP6, LTBP1, CRABP1, SOSTDC1, IL32, PRELP, IDO1, TIMP1, EMILIN1, PDPN, SELM, AKR1C1, C7, AC090498.1, C2, IGFBP5, MMP9, TNC, G0S2, IFI27L2, PTGES, PDGFRA, UBD, S100A13, CLU, ISOC1, FMOD, NDRG2, FRZB, TYMP, ADAMDEC1, TSPAN8, ITIH5, LGALS3BP, CLMP, IL34, CFB, GEM, PPA1, SQSTM1, CD63, TMEM98, CEBPB, DDR2
--	--

By performing ssGSEA on the mRNA expression levels of the characteristic genes for each of the four cell components, the relative abundance of these components in NPC tumor tissues can be predicted.

The expression levels of these cell components are further demonstrated to predict clinical prognosis in NPC patients.

To evaluate the prognostic effect of each component, survival analysis was conducted. Pretreatment tumor tissue samples were collected via nasopharyngeal biopsy from 150 treatment-naïve NPC patients at Sun Yat-sen University Cancer Center. mRNA microarray profiling was performed on each sample to obtain gene expression profiles.

Figure 2 shows the characteristic scores of germinal center B cell components, follicular helper CD4⁺ T cell components, CXCL13⁺ CD8⁺ T cell components, and CXCL13⁺ fibroblast components (rows) in 150 NPC samples (columns). The color gradient ranging from blue to red represents scaled expression levels from low to high. Colored bars at the top of the heatmap indicate overall survival and progression-free survival information. All samples were grouped into two groups based on low- or high-level TLS.

Using ssGSEA, the expression levels of the four cell components were quantified. ROC

curve analysis was applied to determine optimal threshold for distinguishing high- and low-expression groups for each component. Survival analysis was then performed to compare PFS and OS between high- and low-expression groups.

Figure 3 shows the KM survival curve analysis of PFS (left panel) and OS (right panel) differences in NPC patients grouped by high- and low-level TLS.

The result in Figure 4 shows that patients with high expression of any one of the four cell components exhibited significantly better prognosis compared to those with low expression ($P < 0.05$) (Figure 4).

The present invention demonstrates that by quantifying the expression levels of characteristic genes or proteins of the four immune cell components, it is possible to predict the relative abundance of distinct cell components within TLS in NPC and further predict clinical prognosis. This approach facilitates personalized treatment strategies, ultimately improving patient survival outcomes.

For judging high risk vs low risk in risk stratification, the following thresholds can be applied: if the expression level of germinal center B cell component is below 10.77, it is classified as high risk; if the expression level of germinal center B cell component is above 10.77, it is classified as low risk; if the expression level of follicular helper CD4⁺ T cell component is below 9.09, it is classified as high risk; if the expression level of follicular helper CD4⁺ T cell component is above 9.09, it is classified as low risk; if the expression level of CXCL13⁺ CD8⁺ T cell component is below 8.41, it is classified as high risk; if the expression level of CXCL13⁺ CD8⁺ T cell component is above 8.41, it is classified as low risk; and if the expression level of CXCL13⁺ fibroblast component is below 8.58, it is classified as high risk; if the expression level of CXCL13⁺ fibroblast component is above 8.58, it is classified as low risk.

Alternative methods for determining specific risk thresholds may also be applied based on established clinical or analytical protocols.

The above embodiments are provided for illustrative purposes only and do not limit the scope of the present invention in any manner. Those one skilled in the art should understand that modifications or substitutions to the technical details or implementations of the present invention may be made without departing from its spirit and scope, and all such modifications or substitutions shall fall within the protective scope of the present invention.

Claims

1. A combination of tertiary lymphoid structure component marker for predicting prognosis of nasopharyngeal carcinoma, wherein the tertiary lymphoid structure component marker comprises one or more component markers as follows: germinal center B cell component marker, follicular helper CD4⁺ T cell component marker, CXCL13⁺ CD8⁺ T cell component marker, and CXCL13⁺ fibroblast component marker.

2. The combination of tertiary lymphoid structure component marker of claim 1, wherein the germinal center B cell component marker comprises one or more characteristic genes or proteins as follows: RGS13, MARCKSL1, NEIL1, TCL1A, HMGN1, LRMP, ACTG1, SERPINA9, RP11-231C14.7, LPP, CD79B, ATP5L, SERF2, SUGCT, LTB, HMCES, BASP1, LMO2, RFTN1, SMIM14, BCL7A, DAAM1, AC023590.1, BIK, ARPC2, PARP1, GCHFR, UCP2, TCEA1, METAP2, CD40, GAPDH, CD22, BCAS4, SYNE2, LIMD2, VPREB3, CCDC144A, TKT, SLC25A5, CORO1A, HMGA1, IRF8, STAG3, EZR, CFL1, LCP1, ACTB, H3F3A, CD81, CCDC88A, GRHPR, SUSD3, VNN2, ACY3, TMEM123, KLHL6, PRPSAP2, PRDX6, ALOX5AP, GMDS, DHRS9, AC079767.4, PFN1, WDR66, AICDA, GGA2, BRK1, ARPC1B, CD53, ANP32B, MEF2B, ARPC5, DEF8, ACTR3, OAZ1, MYO1E, SWAP70, ARPC4, PKM, SMARCB1, HOPX, HLA-DMB, LCK, ARPC3, SRSF9, HTR3A, RRAS2, MBD4, SEC14L1, IL4R, DCAF12, LYPLA1, SNAP23, LAMTOR5, UBE2J1, CCDC69, HLA-DMA, RGS10, CDV3.

3. The combination of tertiary lymphoid structure component marker of claim 1, wherein the follicular helper CD4⁺ T cell component marker comprises one or more characteristic genes or proteins as follows: CXCL13, ITM2A, RP5-1028K7.2, PDCD1, TOX2, CHI3L2, ICA1, SH2D1A, IGFBP4, PASK, PCAT29, NR3C1, RP11-455F5.5, SMCO4, CD200, NMB, TOX, PTPN13, CXCR5, ZNF331, TCF7, TIGIT, FKBP5, RNF19A, LAT, MAGEH1, C9orf16, LIMS1, FABP5, PPP1CC, SRGN, THADA, CPM, IL6ST, CORO1B, ST8SIA1, COTL1, IKZF3, SLC9A9, RGS2, IL6R, MAF, C1orf228, CXCR4, FYN, GNG4, NFATC1, YPEL5, TBC1D4, JUNB, TRBC2, NUCB2, DUSP6, GZMM, LGMN, RNASET2, TC2N, TSHZ2,

CD84, IL16, SFXN1, TRAC, SIRPG, FYB, LRMP, CD40LG, BTLA, RHOB, POU2AF1, NR4A2, PPP2R5C, AHI1, SH3TC1, TRPS1, DENND2D, ASB2, KIAA1551, RAB27A, STOM, PTPRC, FAAH2, DUSP2, CD27, SIAH2, EVI2B, FAM118A, TSPYL2, SEPTIN6, TNFSF8, ZBTB10, SESN3, PYHIN1, TIFA, PTPN2, SPCS1, CNIH1, TMEM2, SLC25A46, H2AFZ, PTPN7.

4. The combination of tertiary lymphoid structure component marker of claim 1, wherein the CXCL13⁺ CD8⁺ T cell component marker comprises one or more characteristic genes or proteins as follows: CXCL13, TRBV6-5, TOX, CD27, FXYD6, RP5-1028K7.2, GZMK, CD74, TSHZ2, DUSP4, IGFBP4, SH2D1A, FOS, BCAT1, AP3S1, CHI3L2, CAV1, HLA-DPB1, ENC1, GEM, NR3C1, ITM2A, HLA-DPA1, HLA-DQA1, PASK, LANCL2, GZMM, FABP5, CRTAM, PMAIP1, DUSP1, ITM2C, VOPP1, VIPR2, TCF7, CDHR1, RGS2, CD82, HLA-DRA, CLDND1, ZNF331, APLP2, FCRL3, JUNB, TUBA1A, SEPTIN6, TIFA, CHN1, TRAT1, AIG1, HLA-DRB1, TRAV3, HSPA1B, RHOB, TTN, CXCR5, DKK3, TIAM1, PCAT29, DNAJB1, CD200, B3GNT2, FAAH2, CDH1, LY9, EOMES, CD8A, SEPTIN9, PEBP1, SIRPG, ELMO1, HLA-A, PDLIM4, SMPDL3A, LAT, DTHD1, CXCR3, HLA-DRB5, HLA-B, PON2, IGFBP2, STAG3, BATF, GRINA, LYST, CIRBP, IL10RA, AIF1, LPL, METTL8, HIF1A, CD28, PDE4DIP, SEMA4A, ZFAS1, L1CAM, MLLT3, NGFRAP1, TNFRSF9, HLA-DQB1.

5. The combination of tertiary lymphoid structure component marker of claim 1, wherein the CXCL13⁺ fibroblast component marker comprises one or more the characteristic genes or proteins as follows: PTGDS, CCL19, MMP1, MMP3, CXCL14, CXCL1, CXCL13, CHI3L1, CCL11, C3, RARRES2, CCL21, SOD2, TNFSF13B, C1S, TMEM176B, VCAM1, ABI3BP, HPGD, CP, ADH1B, DPT, CD24, CCL2, CXCL6, TMEM176A, MT2A, RBP5, LUM, MFAP4, DCN, CYP1B1, FGF7, FTH1, CXCL12, C1R, DIO2, CTSK, LSAMP, PCOLCE, FDCSP, RND3, CTSS, CXCL9, NR2F1, APOE, RARRES1, LOXL1, FMO2, PLEKHH2, OLFML3, CCDC80, TWIST2, STEAP1, CD82, SERPINF1, TNFAIP6, LTBP1, CRABP1, SOSTDC1, IL32, PRELP, IDO1, TIMP1, EMILIN1, PDPN, SELM, AKR1C1, C7, AC090498.1, C2,

IGFBP5, MMP9, TNC, G0S2, IFI27L2, PTGES, PDGFRA, UBD, S100A13, CLU, ISOC1, FMOD, NDRG2, FRZB, TYMP, ADAMDEC1, TSPAN8, ITIH5, LGALS3BP, CLMP, IL34, CFB, GEM, PPA1, SQSTM1, CD63, TMEM98, CEBPB, DDR2.

6. Use of the combination of tertiary lymphoid structure component marker of any one of claims 1 to 5 in predicting prognosis of nasopharyngeal carcinoma.

7. Use of a product for quantifying the combination of tertiary lymphoid structure component marker in the preparation of the reagent for predicting prognosis of nasopharyngeal carcinoma.

8. The use of claim 7, wherein the product for quantifying the group of tertiary lymphoid structure component marker comprises a reagent for quantifying mRNA or protein levels of the tertiary lymphoid structure component marker.

9. A system for predicting prognosis of nasopharyngeal carcinoma, comprising:

a tertiary lymphoid structure component marker quantification arrangement for determining an expression level of the combination of tertiary lymphoid structure component marker of any one of claims 1 to 5;

a data analysis arrangement for calculating based on the expression level of the combination of tertiary lymphoid structure component marker and determining prognosis of nasopharyngeal carcinoma; and

a result output arrangement for outputting the calculated prognosis.

10. The system of claim 9, wherein the tertiary lymphoid structure component marker quantification arrangement comprises reagents, chips, detection tools, or kits for quantitatively detecting expression levels of mRNA or proteins of the marker.