Supplementary Data 3

To evaluate the performance of the approaches proposed in the manuscript, we performed a comparison of the over-represented GO terms versus PubMed tokens in the HG-U133A-based literature gene lists. Version 2.4.0 of the Bioconductor package GOstats was used to identify GO terms (only the biological process category was considered here) that are significantly enriched in these gene lists. Over-represented tokens were identified with the outlier detection procedure (Outlier) and the extended hypergeometric test-based approach (ExtendedHG). In each case, the numbers of significant GO terms and tokens were recorded. For the token-based analysis, we chose to show results from the method (either Outlier or ExtendedHG) that produced the most significant tokens.

52 gene lists were analysed, of which:

- 5 do not have any hits
- 3 have at least one significant GO terms but not tokens
- 15 have at least one significant token but not GO terms

As shown in Figure 1, there is a good correlation between the number of significant GO terms and the number of significant tokens identified across the remaining 29 gene lists, which are associated with at least one significant GO term and token.

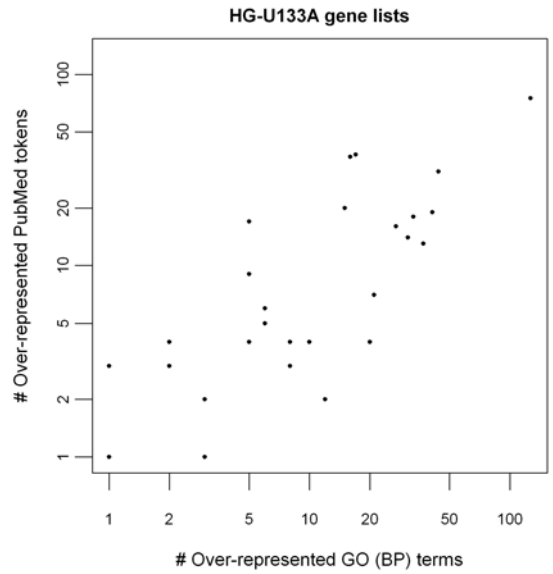

Figure 1. A comparison of the number of significant GO (BP) terms versus the number of significant abstract tokens in the 29 gene lists. The axes are on log scale.

To obtain a 'rough guide' for assessing the relevance of the over-represented GO terms and tokens, we read the publications from which these gene lists were derived and make a note of the biology that is deemed important by the authors. The results are shown in the following table.

ID	Source	Description/biology under studied	Over-represented GO (BP) terms	Over-represented Tokens	[a]Method
hs1b	PMID: 16531451 Title: Global alterations in mRNA polysomal recruitment in a cell model of colorectal cancer progression to metastasis. Extracted from Supplementary Table II.	Probesets changed in the polysomal RNA sample. • Amine metabolism • Amino acid activation and translation • RNA transport and metabolism • cell proliferation • Apoptosis • tRNA aminoacylation pathway	Regulation of progression through cell cycle Regulation of cell cycle	BREAST METASTATIC OVEREXPRESSION	E
hs2a	PMID: 15210650 Title: Gene expression profiles during human CD4+ T cell differentiation. Extracted from Supplementary Table 1.	Transcripts with greater than 3-fold enrichment in every T cell subpopulation compared to TSC. • T cell differentiation and function • Immune response • TCR signalling • Intrathymic differentiation • ERK1/ERK2 activity • Intrathymic T cell selection	Immune system process T cell activation Lymphocyte activation Leukocyte activation Positive regulation of antigen receptor-mediated signaling pathway Hemopoietic or lymphoid organ development Immune system development Cell activation Hemopoiesis Regulation of T cell activation Immune response T cell differentiation Lymphocyte differentiation Regulation of lymphocyte activation Regulation of cell activation	T-CELL LYMPHOID TCR THYMOCYTE CD3 NK LINEAGE HEMATOPOIETIC IL-2 LYMPHOCYTE JURKAT LYMPHOMA IMMUNE LCK T-LYMPHOCYTE CD2 KILLER NAIVE CD45RA B-CELL	E
hs2b	PMID: 15210650 Title: Gene expression profiles during human CD4+ T cell differentiation. Extracted from Supplementary Table 2.	Transcripts whose expression changed by more than 3-fold during T cell differentiation.	Cell cycle Cell cycle process Mitotic cell cycle M phase of mitotic cell cycle Cell cycle phase Mitosis Regulation of cell cycle DNA replication M phase Regulation of progression through cell cycle	LYMPHOID THYMOCYTE ANAPHASE CD8 LYMPHOCYTE B-CELL LYMPHOMA T-CELL MITOSIS SPINDLE	E

ID	Source	Description/biology under studied	Over-represented GO (BP) terms	Over-represented Tokens	[a]Method
			Immune response Cell division Spindle organization and biogenesis Immune system process Programmed cell death Regulation of programmed cell death Apoptosis Regulation of apoptosis Cell death Death DNA metabolic process Cell cycle checkpoint Microtubule-based process Antigen processing and presentation Regulation of mitosis Positive regulation of apoptosis Regulation of biological process Positive regulation of programmed cell death Cell development Negative regulation of biological process Biological regulation Regulation of cellular process Induction of apoptosis Mitotic sister chromatid segregation Negative regulation of cellular process Induction of programmed cell death Sister chromatid segregation Spindle checkpoint Negative regulation of programmed cell death DNA-dependent DNA replication Chromosome segregation Negative regulation of apoptosis Cell differentiation Cellular developmental process	TCR CHECKPOINT CD3 CYTOKINESIS THYMUS CD4 NK IL-2 KINETOCHORE INTERLEUKIN-2 PROMETAPHASE LEUKEMIA INTERFERING MITOTIC INTERPHASE KILLER PROLIFERATING LYMPH MIDBODY LINEAGE VIRUS	
hs2c	PMID: 15210650 Title: Gene expression profiles during human CD4+ T cell differentiation.	Transcripts enriched in both ITTP and DP by more than 3-fold. • DNA replication, recombination and repair • Cell cycle regulation, progression, mitosis	Mitotic cell cycle Mitosis M phase of mitotic cell cycle Cell cycle phase M phase	MITOTIC ANAPHASE SPINDLE MITOSIS KINETOCHORE	O

ID	Source	Description/biology under studied	Over-represented GO (BP) terms	Over-represented Tokens	[a]Method
	Extracted from Supplementary Table 3-1.	• Lipid/glycolipid-presenting CD1 family • Transcriptional regulation • Regulation of apoptosis	Cell cycle Cell cycle process Cell division Spindle organization and biogenesis Microtubule-based process Microtubule cytoskeleton organization and biogenesis Regulation of mitosis Cytoskeleton organization and biogenesis Regulation of progression through cell cycle Regulation of cell cycle Organelle organization and biogenesis Phosphoinositide-mediated signaling Second-messenger-mediated signaling Chromosome segregation DNA metabolic process Spindle checkpoint Microtubule-based movement DNA replication Mitotic sister chromatid segregation Sister chromatid segregation Cytoskeleton-dependent intracellular transport Mitotic spindle organization and biogenesis	CHECKPOINT CONGRESSION CENP-A AURORA BUB1 CENP-H MAD2 CENTROMERE PROMETAPHASE KINETOCHORE-MICROTUBULE MTOC	
hs2d	PMID: 15210650 Title: Gene expression profiles during human CD4+ T cell differentiation. Extracted from Supplementary Table 3-2.	Transcripts enriched in more mature cells (SP4, CB4, and AB4) by more than 3-fold. • Intracellular communication • Cell surface receptors • Peptide-presenting MHC antigens • Transcriptional regulation • Regulation of apoptosis	Antigen processing and presentation of peptide antigen via MHC class I Antigen processing and presentation of peptide antigen Immune response Antigen processing and presentation Immune system process	HLA-CLASS HLA-A OR-C PERIPHERAL CD8 HLA-C HLA-B HLA-G IFN-GAMMA	O
hs2e	PMID: 15210650 Title: Gene expression profiles during human CD4+ T cell differentiation. Extracted from Supplementary Table 3-3.	Transcripts enriched by more than 3-fold in ITTP compared to other lymphocytes. • Immune function	Nitric oxide mediated signal transduction CGMP biosynthetic process CGMP metabolic process	INTERLEUKIN-2	O
hs2f	PMID: 15210650	Transcripts enriched by more than 3-fold in	Lymphocyte activation	LYMPHOPROLIFERATIO	O

ID	Source	Description/biology under studied	Over-represented GO (BP) terms	Over-represented Tokens	[a]Method
	Title: Gene expression profiles during human CD4+ T cell differentiation. Extracted from Supplementary Table 3-4.	DP compared to other lymphocytes. • Thymocyte differentiation • Thymocyte survival and positive selection • Modulation of Th1 and Th2 response • CD1 family proteins • T cell co-stimulation	Leukocyte activation Cell activation T cell activation Antigen processing and presentation	N MYCOBACTERIAL MYCOBACTERIA INTERACTION	
hs2i	PMID: 15210650 Title: Gene expression profiles during human CD4+ T cell differentiation. Extracted from Supplementary Table 4-1.	Transcripts showing SP4>CB4>AB4 pattern. • Plasma membrane proteins	-	HMG-BOX MMP-2	O
hs3a	PMID: 15897907 Title: Identification of molecular apocrine breast tumours by microarray analysis. Extracted from Supplementary Table Sheet 2.	Genes which best discriminate apocrine vs luminal (AL).	Alcohol metabolic process Monocarboxylic acid metabolic process Carboxylic acid metabolic process Organic acid metabolic process Steroid biosynthetic process Lipid metabolic process Sterol biosynthetic process Aldehyde metabolic process Fatty acid metabolic process Cholesterol biosynthetic process	METABOLISM ACETOXYMETHYL ANTIANDROGEN ANDROGEN	O
hs3b	PMID: 15897907 Title: Identification of molecular apocrine breast tumours by microarray analysis. Extracted from Supplementary Table Sheet 2.	Genes which best discriminate apocrine vs basal (AB).	Lipid metabolic process Carboxylic acid metabolic process Organic acid metabolic process Monocarboxylic acid metabolic process Alcohol metabolic process Sterol biosynthetic process Cellular lipid metabolic process Lipid biosynthetic process Steroid biosynthetic process Fatty acid metabolic process Cholesterol biosynthetic process Sterol metabolic process	DESATURASE FATTY	O
hs4a	PMID: 16260967 Title: Effects of aerobic training on gene expression in skeletal muscle of elderly	Genes whose expression increased after training. • Energy metabolism/mitochondrion • Lipid metabolism	Acetyl-CoA metabolic process Tricarboxylic acid cycle intermediate metabolic process Tricarboxylic acid cycle Acetyl-CoA catabolic process	MITOCHONDRIAL MITOCHONDRIA CATALYSIS BURY	O

ID	Source	Description/biology under studied	Over-represented GO (BP) terms	Over-represented Tokens	[a]Method
	men. Extracted from Table S2 in the main paper.	• Proton pumps • Collagen • Protein, amino acid dephosphorylation • Heme biosynthesis	Cofactor metabolic process Coenzyme catabolic process Cellular catabolic process Cofactor catabolic process		
hs4b	PMID: 16260967 Title: Effects of aerobic training on gene expression in skeletal muscle of elderly men. Extracted from Table S3 in the main paper.	Genes whose expression decreased after training. • Ribosome and protein catabolism • Muscle degradation	Translation Macromolecule biosynthetic process Biosynthetic process Cellular protein metabolic process Cellular macromolecule metabolic process Protein metabolic process	RIBOSOMAL S14 R-PROTEIN RRNA U14	O
hs5a	PMID: 12958056 Title: Gene expression profiling of bronchoalveolar lavage cells in acute lung rejection. Extracted from Supplementary Table E1.	Genes that are up-regulated in gene expression in acute rejection vs. no rejection (False Discovery Rate = 0.94%). • Acute rejection response • Immune response • Inflammatory response • Transcriptional regulation • TGF-beta signalling • Apoptosis • Nucleotide GPCR receptors • Peptide GPCR receptors • Wnt signalling • Cytokine-CXC chemokine pathways	Immune system process Immune response Leukocyte activation Response to stimulus Cell activation Lymphocyte activation Positive regulation of lymphocyte activation Regulation of lymphocyte activation Regulation of cell activation Positive regulation of isotype switching to IgG isotypes Lymphocyte mediated immunity Cellular defense response Leukocyte mediated immunity Adaptive immune response Adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains Regulation of immune system process Regulation of immune response Isotype switching to IgG isotypes Regulation of isotype switching to IgG isotypes Immunoglobulin mediated immune response B cell mediated immunity Positive regulation of B cell activation Immune effector process Positive regulation of isotype switching Immune response-activating cell surface receptor signaling pathway	THYMOCYTE CD8 NK TCR CD4 CD3 KILLER IL-2 T-CELL IMMUNE LYMPHOCYTIC LYMPHOCYTE MAB CYTOLYTIC JURKAT LYMPHOPROLIFERATIVE LYMPHOMA B-CELL	E

ID	Source	Description/biology under studied	Over-represented GO (BP) terms	Over-represented Tokens	[a]Method
			Immune response-activating signal transduction Immune response-regulating signal transduction Immune response-regulating cell surface receptor signaling pathway Antigen receptor-mediated signaling pathway Positive regulation of mononuclear cell proliferation Positive regulation of lymphocyte proliferation Regulation of B cell activation Signal transduction		
hs5b	PMID: 12958056 Title: Gene expression profiling of bronchoalveolar lavage cells in acute lung rejection. Extracted from Supplementary Table E2.	Genes with significant changes in gene expression in acute rejection vs. no rejection (False Discovery Rate = 4.63 %). • Acute rejection response • Immune response • Inflammatory response • Transcriptional regulation • TGF-beta signalling • Apoptosis • Nucleotide GPCR receptors • Peptide GPCR receptors • Wnt signalling • Cytokine-CXC chemokine pathways	Immune system process Immune response Response to stimulus Leukocyte activation Lymphocyte activation Cell activation Cell surface receptor linked signal transduction T cell activation Regulation of lymphocyte activation Regulation of cell activation Cellular defense response Signal transduction Positive regulation of lymphocyte activation Cell communication Regulation of T cell activation Defense response Regulation of multicellular organismal process	CD4 NK IL-2 CD3 CD8 KILLER THYMOCYTE ENGAGEMENT TCR T-CELL CYTOLYTIC CD16 CD56 MAB IMMUNE LYMPHOCYTE CD2 IL-12 IMMUNOGLOBULIN PBL INTERLEUKIN ALLOGENEIC LIGATION CTL GRANZYME CYTOTOXICITY RAFT TH1 JURKAT MONOCYTE	E

ID	Source	Description/biology under studied	Over-represented GO (BP) terms	Over-represented Tokens	[a]Method
				EFFECTOR CYTOKINE PHYTOHEMAGGLUTININ ANTI-CD3 PERIPHERAL MONONUCLEAR CD45RO IFN-GAMMA	
hs6a	PMID: 16319128 Title: Distinct expression profile in fumarate-hydratase-deficient uterine fibroids. Extracted from Supplementary Table 1.	Down-regulated genes in FH mutant relative to FH wild-type fibroids. • Extracellular matrix • Cell mobility • Muscle contraction • Organogenesis • Muscle development • Cell adhesion • Plasma membrane	Anatomical structure development	MICROFIBRIL TRANSFORMING SMOOTH	
hs6b	PMID: 16319128 Title: Distinct expression profile in fumarate-hydratase-deficient uterine fibroids. Extracted from Supplementary Table 1.	Up-regulated genes in FH mutant relative to FH wild-type fibroids. Extracted from Supplementary Table 1. • Glycolysis • Carbohydrate metabolism • Hexose metabolism • Iron ion homeostasis • Oxidoreductase activity • Membrane lipid catabolism • Integral to endoplasmic reticulum membrane • Electron transporter activity	Glucose catabolic process Glycolysis Hexose catabolic process Monosaccharide catabolic process Alcohol catabolic process Glucose metabolic process Cellular carbohydrate catabolic process Hexose metabolic process Monosaccharide metabolic process Carbohydrate catabolic process Cellular catabolic process Cellular carbohydrate metabolic process Carbohydrate metabolic process Catabolic process Cellular macromolecule catabolic process Alcohol metabolic process Phospholipid catabolic process Macromolecule catabolic process Cellular iron ion homeostasis Iron ion homeostasis	GLYCOLYTIC ALDO-KETO PEROXIDATION NICOTINAMIDE	O
hs6d	PMID: 16319128	Up-regulated genes in FH mutant relative	Glucose catabolic process	GLYCOLYTIC	O

ID	Source	Description/biology under studied	Over-represented GO (BP) terms	Over-represented Tokens	[a]Method
	Title: Distinct expression profile in fumarate-hydratase-deficient uterine fibroids. Extracted from Supplementary Table 3.	to normal myometrium. • Carbohydrate metabolism • Glycolysis	Hexose catabolic process Monosaccharide catabolic process Alcohol catabolic process Cellular carbohydrate catabolic process Carbohydrate catabolic process Glycolysis Glucose metabolic process Cellular carbohydrate metabolic process Hexose metabolic process Monosaccharide metabolic process Carbohydrate metabolic process Cellular catabolic process Macromolecule catabolic process Alcohol metabolic process Catabolic process Cellular macromolecule catabolic process NADP metabolic process Nicotinamide metabolic process Cellular iron ion homeostasis Iron ion homeostasis	NONSPHEROCYTIC NADP HEMOLYSIS APOFERRITIN RESOLUTION ISOENZYME	
hs8	PMID: 15971941 Title: Derivation of multipotent mesenchymal precursors from human embryonic stem cells. Extracted from Supplementary Table S2.	Genes shared between primary and hESC-derived mesenchymal precursors but significantly different from undifferentiated hESCs. • MSC markers including mesenchymal stem cell protein DSC54, hepatocyte growth factor, neuropilin I, forkhead box D1, notch homolog.	Organ development Hemostasis Cell cycle arrest Blood coagulation Coagulation Wound healing Response to wounding Regulation of body fluids	COLLAGEN ECM STROMAL	O
hs11a	PMID: 16049480 Title: Genes that mediate breast cancer metastasis to lung. Extracted from Supplementary Table 2.	Genes differentially expressed between parental MDA-MB-231 and LM2 cell lines selected to be highly metastatic to lung. • Lung metastatic activity • Growth and survival factors • Chemokines • Cell adhesion receptors • Extracellular proteases • Intracellular enzymes • Transcriptional regulators	Antigen processing and presentation of peptide or polysaccharide antigen via MHC class II Immune response Immune system process	METALLOPROTEINASE-2 SBT	O

ID	Source	Description/biology under studied	Over-represented GO (BP) terms	Over-represented Tokens	[a]Method
hs11b	PMID: 16049480 Title: Genes that mediate breast cancer metastasis to lung. Extracted from Supplementary Table 4.	Lung metastasis candidate genes. • Extracellular proteins (eg. SPARC, MMP2) act as virulence genes that may allow tumours to invade, colonise and grow in the lungs.	-	METALLOPROTEINASE-2	O
hs12b	PMID: 16089502 Title: Functional analysis of human hematopoietic stem cell gene expression using zebrafish. Extracted from Supplementary Table S2.	Probesets differentially expressed between adult bone marrow derived Rho-lo and Rho-hi cells. • Cell cycle control	Cell cycle Mitotic cell cycle Cell cycle process Cell cycle phase DNA metabolic process M phase of mitotic cell cycle Mitosis M phase DNA replication Cellular metabolic process Cell division Primary metabolic process Metabolic process Regulation of cell cycle Macromolecule metabolic process Regulation of progression through cell cycle Chromosome organization and biogenesis Translation Chromosome organization and biogenesis (sensu Eukaryota) Cell cycle checkpoint Chromosome segregation Response to DNA damage stimulus DNA-dependent DNA replication Response to endogenous stimulus Nucleosome assembly Mitotic sister chromatid segregation Spindle organization and biogenesis Interphase Macromolecule biosynthetic process Sister chromatid segregation DNA repair Organelle organization and biogenesis	ANAPHASE CHECKPOINT MITOSIS MITOTIC RIBOSOMAL CYTOKINESIS INTERPHASE CYCLE G1 KINETOCHORE REPLICATION CDK2 PROLIFERATING	E

ID	Source	Description/biology under studied	Over-represented GO (BP) terms	Over-represented Tokens	[a]Method
			Biopolymer metabolic process Chromatin assembly Biosynthetic process Mitotic spindle organization and biogenesis Nucleobase, nucleoside, nucleotide and nucleic acid metabolic process		
hs12c	PMID: 16089502 Title: Functional analysis of human hematopoietic stem cell gene expression using zebrafish. Extracted from Supplementary Table S3.	Probesets differentially expressed between Rho-lo and Rho-hi cells from both umbilical cord blood and adult bone marrow. • Hematopoietic differentiation and development • Cell cycle control • Wnt signalling • Germ cell development • Globins	Oxygen transport Gas transport Antigen processing and presentation of peptide or polysaccharide antigen via MHC class II DNA replication Antigen processing and presentation DNA metabolic process	THAL BETA-CHAIN DELTA-GLOBIN HBA2 THALASSEMIA ANODE	O
hs15a	PMID: 12756304 Title: A global view of the selectivity of zinc deprivation and excess on genes expressed in human THP-1 mononuclear cells. Extracted from Supplementary Table 3.	Group 1 zinc responsive genes. • Nucleic acid binding • Apoptosis • Metabolism • Cell growth and development • Signal transduction • Immune, cytokine • Cytoskeleton	RNA metabolic process Transcription from RNA polymerase II promoter	SCLEROTOME CD4 RNAP EFFECTOR	
hs17b	PMID: 16804116 Title: Gene-expression profiling of Waldenstrom macroglobulinemia reveals a phenotype more similar to chronic lymphocytic leukemia than multiple myeloma. Extracted from Supplementary Table S1: MM Unique Genes.	Genes that displayed distinct expression profile in MM compared to CLL and WM. • Signal transduction and intracellular signalling • Cell-surface receptor-linked signalling e.g. AKT, IGF-1R and Wnt signalling • Prostacyclin synthesis • angiopoientin signalling • Integrin-mediated cell adhesion • Early B-cell receptor signalling	Regulation of biological process Regulation of cellular process Biological regulation	MB-1	O
hs17c	PMID: 16804116 Title: Gene-expression profiling of Waldenstrom macroglobulinemia reveals a phenotype more similar to chronic lymphocytic leukemia than multiple myeloma.	Genes that displayed distinct expression profile in CLL compared to WM and MM. • Apoptosis regulation • Immune response • Cell cycle regulation	-	APC	O

ID	Source	Description/biology under studied	Over-represented GO (BP) terms	Over-represented Tokens	[a]Method
	Extracted from Supplementary Table S1: CLL Unique Genes.				
hs17d	PMID: 16804116 Title: Gene-expression profiling of Waldenstrom macroglobulinemia reveals a phenotype more similar to chronic lymphocytic leukemia than multiple myeloma. Extracted from Supplementary Table S1: WM CLL B-cell cluster.	A cluster of genes that were over-expressed in B-cell, WM and CLL. • Cell cycle regulation	Immune system process Immune response Cell communication Signal transduction Lymphocyte activation Leukocyte activation Cell activation B cell activation Defense response Immune response-activating cell surface receptor signaling pathway Immune response-activating signal transduction Immune response-regulating signal transduction Immune response-regulating cell surface receptor signaling pathway Antigen receptor-mediated signaling pathway Immune system development T cell activation	LYMPHOID B-CELL HEMATOPOIETIC LYMPHOCYTE CD19 ENGAGEMENT LINEAGE PRE-B CD8 TCR LYMPHOMA NAIVE BCR CD4 THYMOCYTE SRC IMMUNE MONOCYTE PRO-B BURKITT IL-4 CD21 GERMINAL TONSILLAR CD23 LYMPHOCYTIC IMMUNOGLOBULIN CD45 CD3 B-LYMPHOCYTE EXTRANODAL HISTOCOMPATIBILITY F-ACTIN JURKAT ZAP70 NK	E

ID	Source	Description/biology under studied	Over-represented GO (BP) terms	Over-represented Tokens	[a]Method
				KILLER	
hs18a	PMID: 16836768 Title: Signatures of human regulatory T cells: an encounter with old friends and new players. Extracted from Additional file 1.	Up-regulated genes comparing CD4+CD25+ T cells versus CD4+CD25- T cells. All differentially expressed genes can be classified into: • Cytokines/chemokines and their receptors • Cell cycle and proliferation • Apoptosis • Signal transduction • Transcriptional regulation The authors identified 3 signalling modules using Pathway Analysis software: • Genes that control T cell receptor signalling, activation and proliferation • Genes that control differentiation and maintenance • Genes that control survival/apoptosis	Antigen processing and presentation of peptide or polysaccharide antigen via MHC class II Antigen processing and presentation Immune response Response to stimulus Immune system process	DPB1 DRB1 DRB DR2 HLA-DPB1 DPA1 DQB1*0602 HLA-DRB1 DQW1 DQB1*0302 AND-DQ PCR-SSOP HLA-DR HLA-D DQA1 OLIGOTYPING SBT	O
hs18c	PMID: 16836768 Title: Signatures of human regulatory T cells: an encounter with old friends and new players. Extracted from Additional file 4.	Genes differentially expressed in Foxp3 over-expressing CD4+ Th cell lines cells relative to the GFP transduced CD4+ Th controls. • TNF receptor superfamily • Activation of signal transduction pathways eg. NFkB, JNK, P38, ERK and PI3K • Immune response	Immune system process Immune response Response to stimulus Signal transduction Cell communication Lymphocyte activation Leukocyte activation T cell activation Cell death Death Biological regulation Apoptosis Programmed cell death Cell activation Defense response Cell development Regulation of cellular process Regulation of biological process Positive regulation of biological process Regulation of lymphocyte activation Positive regulation of cellular process	LYMPHOID T-CELL CD4 CD3 CD8 IL-2 CD25 ANTI-CD3 NAIVE LYMPHOCYTE TCR HELPER CD28 ENGAGEMENT IMMUNODEFICIENCY FOXP3 TH1 JURKAT INFECT NK THYMOCYTE	E

ID	Source	Description/biology under studied	Over-represented GO (BP) terms	Over-represented Tokens	[a]Method
			Regulation of cell activation Cell differentiation Cellular developmental process Cell surface receptor linked signal transduction Regulation of apoptosis Regulation of programmed cell death Negative regulation of biological process Negative regulation of cellular process Positive regulation of lymphocyte activation Elevation of cytosolic calcium ion concentration Cytosolic calcium ion homeostasis Cytokine biosynthetic process Cytokine metabolic process Developmental process Regulation of cytokine biosynthetic process Cell proliferation Regulation of T cell activation Cellular di-, tri-valent inorganic cation homeostasis Di-, tri-valent inorganic cation homeostasis Cellular cation homeostasis Cation homeostasis Cellular calcium ion homeostasis Calcium ion homeostasis Inflammatory response Response to external stimulus Cellular ion homeostasis Cellular chemical homeostasis Somatic recombination of immunoglobulin genes during immune response Somatic diversification of immunoglobulins during immune response Immunoglobulin production during immune response Somatic diversification of immune receptors via germline recombination within a single locus Somatic cell DNA recombination Somatic recombination of immunoglobulin gene segments Isotype switching Cellular metal ion homeostasis	IMMUNE PBMC REJECTION B-CELL IL-10 LYMPHOCYTIC UNINFECT KILLER COSTIMULATORY IMMUNITY MONOCYTE LYMPHOMA HIV-1 T-LYMPHOCYTE IL-4 INTERLEUKIN-2 CYTOMETRY INTERLEUKIN HIV ALLOGENEIC CYTOKINE CD45RO INFECTION VIRAL PHA CTLA-4 PHYTOHEMAGGLUTININ GRAFT-VERSUS-HOST IFN-GAMMA HLA-DR PROLIFERATE MAB IL-6 MONONUCLEAR LIGATION CD56 TREG AB INTERFERON-GAMMA	

ID	Source	Description/biology under studied	Over-represented GO (BP) terms	Over-represented Tokens	[a]Method
			Metal ion homeostasis Chemical homeostasis Response to wounding Positive regulation of protein biosynthetic process Regulation of isotype switching Immune system development Ion homeostasis Somatic diversification of immune receptors Somatic diversification of immunoglobulins Production of molecular mediator of immune response Leukocyte differentiation Positive regulation of protein metabolic process Mononuclear cell proliferation Lymphocyte proliferation Cytokine production Positive regulation of cellular biosynthetic process Positive regulation of T cell activation Anti-apoptosis Positive regulation of isotype switching Immunoglobulin production Antigen processing and presentation of peptide or polysaccharide antigen via MHC class II Lymphocyte mediated immunity Regulation of cell proliferation T cell differentiation Negative regulation of apoptosis Cellular defense response Negative regulation of programmed cell death Chemotaxis Taxis Leukocyte mediated immunity Homeostatic process Lymphocyte differentiation Positive regulation of biosynthetic process Adaptive immune response Adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains Regulation of mononuclear cell proliferation	GVHD INFILTRATION ANTI STIMULATION COSTIMULATION IL-7 INFECTIOUS PMA IL-13 CYTOTOXIC THYMIC CD69 IL-12 IMMUNOLOGICAL TNFR	

ID	Source	Description/biology under studied	Over-represented GO (BP) terms	Over-represented Tokens	[a]Method
			Regulation of lymphocyte proliferation Regulation of multicellular organismal process Positive regulation of cytokine biosynthetic process Cytokine and chemokine mediated signaling pathway Locomotory behavior Cellular homeostasis Regulation of T cell proliferation Regulation of biological quality Behavior Hemopoietic or lymphoid organ development Immune effector process Positive regulation of apoptosis Response to chemical stimulus Positive regulation of cell proliferation Positive regulation of programmed cell death Positive regulation of cellular metabolic process Regulation of translation Induction of apoptosis Induction of programmed cell death B cell proliferation Regulation of interleukin-2 biosynthetic process Immunoglobulin mediated immune response B cell mediated immunity T cell proliferation Response to stress Positive regulation of metabolic process Regulation of protein metabolic process Regulation of B cell activation Cleavage of lamin Cleavage of cytoskeletal proteins during apoptosis Hemopoiesis Regulation of cellular biosynthetic process B cell activation Interleukin-2 production Interleukin-2 biosynthetic process		
hs19b	PMID: 15869706 Title: Clinical and biological characteristics of cervical neoplasias with	Down-regulated genes in FGFR3 mutated tumours relative to FGFR3 wild type tumours.	Immune response Immune system process Response to stimulus Defense response	IFN-GAMMA MONOCYTE IMMUNE KILLER	E

ID	Source	Description/biology under studied	Over-represented GO (BP) terms	Over-represented Tokens	[a]Method
	FGFR3 mutation. Extracted from Additional file 2: Negative Significant Genes.	• Transcriptional regulation	Response to wounding Antigen processing and presentation Inflammatory response Response to external stimulus Antigen processing and presentation of peptide or polysaccharide antigen via MHC class II Activation of immune response Positive regulation of immune system process Positive regulation of immune response Innate immune response Cell adhesion Biological adhesion Regulation of immune system process Regulation of immune response Positive regulation of multicellular organismal process Immune effector process Humoral immune response Chemotaxis Taxis Leukocyte mediated immunity Adaptive immune response Adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains Activation of plasma proteins during acute inflammatory response Complement activation Leukocyte activation Locomotory behavior Regulation of multicellular organismal process Cell activation Lymphocyte mediated immunity Behavior Cell motility Localization of cell Prostaglandin biosynthetic process Prostanoid biosynthetic process Immunoglobulin mediated immune response B cell mediated immunity	HISTOCOMPATIBILITY CD8 INFLAMMATORY CD3 NK MHC INFLAMMATION CHEMOTACTIC RHEUMATOID DECIDUAL CYTOKINE IMMUNITY HLA-DR TNF-ALPHA CD4	

ID	Source	Description/biology under studied	Over-represented GO (BP) terms	Over-represented Tokens	[a]Method
			Acute inflammatory response Response to stress		
hs20	PMID: 15604246 Title: Androgen-induced differentiation and tumorigenicity of human prostate epithelial cells. Extracted from Supplementary Table 1.	Genes differentially expressed between LHSR and LHS. • Androgen receptor signaling	Mitotic cell cycle Cell cycle phase Cell cycle process Cell cycle M phase M phase of mitotic cell cycle Mitosis Cell division Regulation of progression through cell cycle Regulation of cell cycle Chromosome segregation DNA replication Regulation of mitosis Mitotic sister chromatid segregation Regulation of cellular process Sister chromatid segregation Cell cycle checkpoint Spindle organization and biogenesis Regulation of biological process Microtubule-based process Biopolymer metabolic process Organelle organization and biogenesis Cell proliferation Interphase of mitotic cell cycle DNA metabolic process Microtubule cytoskeleton organization and biogenesis Cellular component organization and biogenesis Interphase Mitotic cell cycle checkpoint Cytoskeleton organization and biogenesis Nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	MITOTIC CHECKPOINT SPINDLE ANAPHASE MITOSIS ARREST MICROARRAY CYCLE PROMETAPHASE G2 CHROMATID KINETOCHORE CYTOKINESIS INTERPHASE	E

[a]Method: Text-based ORA method (O = Outlier detection analysis; E = Extended hypergeometric test) that produced the results shown in 'Over-represented Tokens'.