Cell type-specific decomposition of gingival tissue transcriptomes

Momen-Heravi, Fatemeh; Friedman, Richard; Albeshri, Sultan; Sawle, Ashley; Kebschull, Moritz; Kuhn, Alexander; Papapanou, Panos N.

DOI:
10.1177/0022034520979614

Journal of
Dental Research

Cell type-specific decomposition of gingival tissue transcriptomes

Journal:	*Journal of Dental Research*
Manuscript ID	JDR-20-1055.R2
Manuscript Type:	Research Reports
Date Submitted by the Author:	13-Nov-2020
Complete List of Authors:	Momen Heravi, Fatemeh; Columbia University College of Dental Medicine, Periodontics / Oral, Diagnostic & Rehabilitation Sciences Friedman, Richard; Columbia University, Biomedical Informatics Shared Resource, Herbert Irving Comprehensive Cancer Center Albeshri, Sultan; Columbia University College of Dental Medicine, Periodontics / Oral, Diagnostic & Rehabilitation Sciences Sawle, Ashley; University of Cambridge, Cancer Research UK Cambridge Institute Kebschull, Moritz; University of Birmingham Kuhn, Alexandre; HES-SO University of Applied Sciences and Arts, Western Switzerland, Institute of Life Technologies, School of Engineering Papapanou, Panos; Columbia University College of Dental Medicine, Periodontics / Oral, Diagnostic & Rehabilitation Sciences
Keywords:	Gene expression, Periodontal disease(s)/periodontitis, Cell biology, Dental informatics/bioinformatics, Epithelia, Immunity
Abstract:	Genome-wide transcriptomic analyses in whole tissues reflect the aggregate gene expression in heterogeneous cell populations comprising resident and migratory cells, and are unable to identify cell type-specific information. We used a computational method (Population-Specific Expression Analysis; PSEA) to decompose gene expression in gingival tissues into cell type-specific signatures for eight cell types (epithelial cells, fibroblasts, endothelial cells, neutrophils, monocytes/macrophages, plasma cells, T cells and B cells). We used a gene expression dataset generated using microarrays from 120 persons (310 tissue samples; 241 periodontitis-affected and 69 healthy). Decomposition of the whole tissue transcriptomes identified differentially expressed genes in each of the cell types, which mapped to biologically relevant pathways including dysregulation of Th17 cell differentiation, AGE-RAGE signaling, and epithelial mesenchymal transition in epithelial cells. We validated selected PSEA-predicted, differentially expressed genes in purified gingival epithelial cells and B cells from an unrelated cohort (n=15 persons), each of whom contributed with one periodontitis-affected and one healthy gingival tissue sample. Differential expression of these genes by qRT-PCR corroborated the PSEA predictions and pointed to dysregulation of biologically important pathways in periodontitis. Collectively, our results demonstrate the robustness of the PSEA in the

	decomposition of gingival tissue transcriptomes, and its ability to identify differentially regulated transcripts in particular cellular constituents. These genes may serve as candidates for further investigation with respect to their roles in the pathogenesis of periodontitis.

Cell type-specific decomposition of gingival tissue transcriptomes

Fatemeh Momen-Heravi[1*], Richard A. Friedman[2*], Sultan Albeshri[1], Ashley Sawle[3],

Moritz Kebschull[1,4], Alexandre Kuhn[5], Panos N. Papapanou[1#]

[1] Division of Periodontics, Section of Oral, Diagnostic and Rehabilitation Sciences,

College of Dental Medicine; New York, NY, USA

[2] Biomedical Informatics Shared Resource, Herbert Irving Comprehensive Cancer Center and Department of Biomedical Informatics, Vagelos College of Physicians and Surgeons; Columbia University, New York, NY, USA

[3]Cancer Research UK Cambridge Institute, University of Cambridge, Cambridge, UK

[4]School of Dentistry, Institute of Clinical Sciences, University of Birmingham, Birmingham, UK

[5]Institute of Life Technologies, School of Engineering, HES-SO University of Applied Sciences and Arts, Western Switzerland, Switzerland

*Equal contribution

[#]Corresponding author: Dr. Panos N. Papapanou; Division of Periodontics; Section of Oral, Diagnostic and Rehabilitation Sciences, Columbia University College of Dental Medicine, 630 West 168th Street, PH-7E-110
New York, NY, USA 10032. Email: pp192@cumc.columbia.edu

Abstract word count: 225
Total word count: 3,148
Total number of figure/tables: 5
Number of references: 40

ABSTRACT

Genome-wide transcriptomic analyses in whole tissues reflect the aggregate gene expression in heterogeneous cell populations comprising resident and migratory cells, and are unable to identify cell type-specific information. We used a computational method (Population-Specific Expression Analysis; PSEA) to decompose gene expression in gingival tissues into cell type-specific signatures for eight cell types (epithelial cells, fibroblasts, endothelial cells, neutrophils, monocytes/macrophages, plasma cells, T cells and B cells). We used a gene expression dataset generated using microarrays from 120 persons (310 tissue samples; 241 periodontitis-affected and 69 healthy). Decomposition of the whole tissue transcriptomes identified differentially expressed genes in each of the cell types, which mapped to biologically relevant pathways including dysregulation of Th17 cell differentiation, AGE-RAGE signaling, and epithelial mesenchymal transition in epithelial cells. We validated selected PSEA-predicted, differentially expressed genes in purified gingival epithelial cells and B cells from an unrelated cohort (n=15 persons), each of whom contributed with one periodontitis-affected and one healthy gingival tissue sample. Differential expression of these genes by qRT-PCR corroborated the PSEA predictions and pointed to dysregulation of biologically important pathways in periodontitis. Collectively, our results demonstrate the robustness of the PSEA in the decomposition of gingival tissue transcriptomes, and its ability to identify differentially regulated transcripts in particular cellular constituents. These genes may serve as candidates for further investigation with respect to their roles in the pathogenesis of periodontitis.

Key words: gene expression, periodontitis, pathobiology, validation, qRT-PCR, epithelial cells, B cells

INTRODUCTION

Periodontitis is a chronic inflammatory disease that is associated with microbial dysbiosis and characterized by loss of connective tissue attachment and alveolar bone (Kinane et al. 2017). Although our understanding of the pathobiology of the disease has been significantly enhanced in the past two decades, the mapping of intra- and intercellular signaling pathways orchestrating the host response to bacterial dysbiosis is a work in progress (Ebersole et al. 2013; Cekici et al. 2014). Delineation of transcriptomic signatures in the gingival tissues at various disease stages has the potential to elucidate key molecular mechanisms underlying the initiation and progression of periodontitis (Demmer et al. 2008; Sawle et al. 2016). Genome-wide transcriptomic analyses in gingival tissues using microarray and RNA sequencing have been used to this end by our group and others (Demmer et al. 2008; Kebschull et al. 2013; Horie et al. 2016), reflecting the average level of gene expression in a mixed population of cells including resident tissue components (e.g. epithelial cells, fibroblasts) as well as migratory cells responsible for immune surveillance and inflammatory responses (including polymorphonuclear neutrophils, monocytes/macrophages, T and B cells, among others). However, the onset and progression of periodontitis is likely orchestrated by contributions of distinct cell populations whose interactions form complex molecular networks that underlie homeostatic or catabolic processes in the gingival tissues (Takayanagi 2005; Cekici et al. 2014). Analyses based on whole tissue transcriptomes are not suited to identify cell type-specific information, because the expression of a particular gene can increase during the transition from health to disease in one cell type and decrease in another, but these opposing changes will remain largely undetected when only a net change is assessed (Heath et al. 2016). Another shortcoming of the assessment of aggregate fold changes of

expression in mixed cell populations is that they cannot distinguish between true changes in expression and those resulting from dynamic fluctuations in the relative proportion of individual cell types which commonly occur during the transition to a pathological state (Newman et al. 2015).

To mitigate the shortcomings associated with whole tissue transcriptomic signatures, a novel computational method was developed to decompose aggregate gene expression profiles in tissue samples that comprise a heterogeneous cellular composition (Kuhn et al. 2011). The method, termed Population-Specific Expression Analysis (PSEA), uses cell population-specific marker genes to generate individual population expression profiles *in silico* without a need for additional experimental steps such as fluorescence-activated cell sorting or laser-capture microdissection. The method can also account for expression changes that occur due to the differential abundance of particular cells types in each sample (Kuhn et al. 2011). In the present study, we first applied PSEA to transcriptomic datasets derived from gingival tissues harvested from states of gingival health or established periodontitis and identified genes that are differentially expressed specifically in each of eight cell types which represent major constituents of the gingival tissues. Next, we used tissue dissociation and immune-magnetic bead purification methods to isolate epithelial cells and B cells from an independent set of gingival biopsies that were not involved in the above computational decomposition and performed quantitative reverse-transcription polymerase chain reaction (qRT-PCR) assays to validate the PSEA-predicted differential expression of selected genes in states of health and periodontitis. Our results demonstrate that PSEA can be successfully used in the decomposition

of the human gingival transcriptome and can facilitate an understanding of the cell-specific molecular processes that occur in the gingival tissues during the course of periodontitis.

METHODS

PSEA analysis

Gene expression data set

We used a gene expression dataset in gingival tissues that we generated earlier using Affymetrix HG-U133Plus 2.0 microarrays (Papapanou et al. 2009), available through GSE16134. A detailed description of the dataset is presented in Supplemental Methods.

The PSEA method

Decomposition of the gene expression signal into molecular subtypes was performed using the PSEA method (Kuhn et al. 2011). Briefly, probesets expressed in a given cell type were identified by assessing the linear dependence of their expression against the expression of marker probesets uniquely expressed in the particular cell type, among all cell types considered. Differential expression was identified by comparing the slope of healthy tissue samples with that of periodontitis-affected samples. For each given gene (probeset), a $\log_2$ fold change, a p-value of expression in the cell type, and a p-value for differential expression in the cell type were computed. Note that since the expression of each gene tested was plotted against that of the marker genes in the same sample, the number of cells of a given type was constant at each point, and differences in cell composition between health and disease did not confound the analyses.

Identification of marker genes

Marker genes were identified using Gene Expression Barcode 3.0 (McCall et al. 2014) that reports the probability that a given probeset is expressed in a particular cell type (see Supplemental

Methods). To define markers, we identified probesets which had a probability of expression of 1 in a given cell type of interest and of 0 in each of the other cell types under consideration. These genes were derived by set-theoretic (Venn) operations on Barcode entries for each of the following eight cell types: epithelial cells, endothelial cells ($CD31^+$), fibroblasts, monocytes ($CD14^+$), neutrophils, plasma cells, T cells ($CD3^+$), and B cells ($CD19^+$). Additional filtering steps of the marker probesets are described in the Supplemental Methods. The final list of the marker probesets used in the analyses is presented in Supplementary Table 1.

Model fitting

The expression of the 54,675 probesets in all patients was fitted to each of 1,208 linear models, each of which had one or more cell types expressed, but only one cell type differentially expressed. The best model for each probeset was determined using the Akaike information criterion (AIC) units (Akaike 1974). Models within 2 AIC units of the best one were selected. Additional filtering of probesets resulted in their removal from differential expression analyses (Supplemental Methods).

Gene set over-representation analysis

Differentially expressed genes for each cell-type with p-value <0.05 were included to identify potentially over-represented processes according to KEGG and Wiki (WP) pathways, using overrepresentation analysis as implemented in gProfiler (Reimand et al. 2007). Pathways with false discovery rate (FDR)≤0.1 were reported.

Validation of PSEA predicted genes

Validation of PSEA-predicted genes was performed in an independent set of gingival tissue samples (15 pairs of healthy and periodontitis-affected sites). After preparation of single cell suspensions and immunomagnetic separation of epithelial cells and B cells, selected PSEA-predicted genes were validated using a quantitative reverse-transcription polymerase chain reaction (qRT-PCR) method. Experimental details are presented in Supplemental Methods.

RESULTS

PSEA analysis

Table 1 lists PSEA-decomposed, cell type-specific gene expression profiles that fulfilled all filtering steps mentioned above. These included 11 transcripts in epithelial cells, 4 in fibroblasts, 5 in endothelial cells, 13 in neutrophils, 6 in plasma cells, and 8 in B cells. No PSEA-decomposed transcripts fulfilled both the absolute log2 fold change differential expression of >0.4 or the CC=1 filtering steps in monocytes/macrophages or in T cells. A more extensive list of PSEA-decomposed genes by cell type, irrespective of |log2 FC| or CC is presented in Supplementary Table 3. This list includes 29 transcripts in epithelial cells, 12 in fibroblasts, 13 in endothelial cells, 21 in neutrophils, 5 in monocytes/macrophages, 11 in plasma cells, 3 in T cells, and 13 in B cells.

Gene set over-representation analysis

Significantly (FDR≤0.1) enriched KEGG and WP pathways by cell type, based on ≥2 PSEA-decomposed differentially expressed genes, are presented in Table 2. Observe that the input in these analyses included all probesets listed in Supplementary Table 3. Five KEGG pathways (Th17 Cell Differentiation, AGE-RAGE Signaling pathway in Diabetes Complications, Relaxin Signaling pathway, Pathogenetic *Escherichia coli* Infection, and Inflammatory Bowel Disease) and four WP pathways (among which Epithelial to Mesenchymal Transition in Colorectal Cancer, and Cannabinoid Receptor Signaling) were identified as differentially enriched in epithelial cells in periodontitis-affected versus healthy gingival tissues (Fig.1). VEGFA-VEGFR2 signaling and Senescence and Autophagy in Cancer were among the significantly enriched pathways in fibroblasts. Fat Digestion and Absorption (KEGG) and Sterol Regulatory Element-Binding

Proteins (SREBP) signaling (WP) were significantly enriched in endothelial cells, and nuclear receptors meta-pathway (WP) and human complement system pathway (WP) were among those enriched in neutrophils. The Protein Processing in Endoplasmic Reticulum pathway (KEGG) was among those over-represented in plasma cells, and the Human Cytomegalovirus Infection pathway (KEGG) was among those enriched in B cells.

Experimental validation of top PSEA assignments in gingival epithelial cells and B cells

We selected two predicted differentially expressed transcripts in epithelial cells (TGF-β and RORA) and B cells (CAMSAP1 and CERS3) for experimental validation based on high differential fold change, high level of confidence, and presence of translated protein. The qRT-PCR analyses showed statistically significant lower expression levels of both TGFβ-1 and RORA in epithelial cells isolated from periodontitis-affected versus healthy gingival tissues, consistent with the PSEA prediction ($p<0.05$, for both matched and unmatched analysis; Fig.2A-D). Validation in B cells, involved only unmatched samples, as no pairs of healthy/periodontitis-affected samples from the same donor and of sufficient quality were available. We detected a significantly lower expression of CERS3 in B cells isolated from periodontitis-affected sites compared to healthy sites, as predicted ($p<0.05$; Fig.3A), however, no statistically significant difference in the expression of CAMSAP1 could be detected. (Fig.3B).

DISCUSSION

Deciphering molecular signatures that distinguish between healthy gingival tissues and those at different stages of periodontitis offers insights into the pathophysiology of the disease process and may, ultimately, identify potential therapeutic targets. However, detection of cellular-level perturbations based on whole-tissue transcriptomic analyses is challenging due to tissue heterogeneity and cellular population shifts during the transition from health to disease. To the best of our knowledge, we applied for the first time a computational method, PSEA, to decompose whole gingival tissue transcriptomes into cell type-specific differential gene expression between periodontal health and periodontitis. Subsequently, we validated the PSEA computations by assessing the differential expression of specific genes in purified gingival epithelial cells and B cells derived from unrelated healthy- or periodontitis-affected tissue samples using qRT-PCR. Our findings point to the utility of PSEA as an alternative to more labor-intensive and costly methodologies in transcriptomic studies of the pathobiology of periodontitis.

In recent years, several medium- or high-throughput technologies have been introduced to study specific cellular components in heterogeneous tissue samples including single-cell and population-specific transcriptome analysis using qRT-PCR, RNA fluorescence *in situ* hybridization (RNA-FISH), RNA-seq, cDNA microarrays, and serial analysis of gene expression (SAGE) (Hunt-Newbury et al. 2007; Esumi et al. 2008; Deng et al. 2014). Alternatively, use of a computational method such as PSEA can help to decompose these aggregate signals into cell type-specific signatures while partly circumventing a number of technical difficulties associated with the above methodologies. However, a number of limitations associated with our study must be acknowledged. First, PSEA is inherently dependent on availability of cell markers previously

identified; thus, potential inaccuracies in the specificity of the utilized markers may inevitably affect the metadata generated. It is conceivable that the relatively limited number of the available cell type-specific markers, in combination with the diversity of the composite cell populations in the gingival tissues, have limited our power to detect differentially expressed transcripts of low abundance, or to accurately predict transcription in less populous cell types. Discovery of additional cell-specific markers and use of larger databases may address these shortcomings in future work. However, we emphasize that PSEA decomposition does not require consideration of every conceivable cell type in the tissues, provided that the models generated using the ultimately filtered probes have a good statistical fit (i.e., high R^2 values), as was the case in our analyses (Supplemental Table 3). Thus, the fact that we did not include less abundant cell types that occur in the gingiva in our models did not affect our inferences regarding genes differentially expressed in the eight studied cell types.

In our validation experiments, we selected two PSEA-predicted differentially expressed genes, on the basis of maximum absolute fold change and high confidence coefficient, in two cell types that are highly prevalent in periodontal tissues (epithelial cells and B-cells), and used pairs of healthy- and periodontitis-affected gingival tissues from 15 *de novo* recruited individuals. As the cells of interest were dissociated and cryopreserved immediately, the distortion in the transcriptional profiles after tissue harvesting was kept to a minimum, as recently demonstrated in a comprehensive study (Guillaumet-Adkins et al. 2017). The amount of tissue harvested in each biopsy did not allow us to separate additional cell types, and the RNA obtained from each cell subset did not allow us to validate more than two genes in each. Thus, these experiments should be viewed as a "proof of principle" validation of the PSEA method in the context of gingival

transcriptomes, rather than as specific verification of each predicted probe. Additional validation studies will be obviously necessary for additional specific cell types and genes of interest.

The PSEA-predicted lower expression of TGF-β in epithelial cells in periodontitis-affected tissues was validated in purified epithelial cells from independent samples. Epithelial cells are the first line of defense against toxic stimuli and periodontal pathogens, orchestrate oral tissue homeostasis, and play crucial roles in the initiation of dysbiotic changes at the dento-gingival niche (Cekici et al. 2014). Epithelial cell-derived TGF-β plays a pivotal role in maintaining a balance between tolerance and immunity (Denney et al. 2015) and exerts its functions through activation of intracellular Smad2/3 proteins and suppression of inflammatory pathways. Concomitantly, TGF-β promotes expression of adhesion molecules and tight junction proteins such as Claudin-1 which maintain epithelial barrier integrity (Howe et al. 2005). In intestinal epithelia, TGF-β is a potent inducer of epithelial cell margination, an essential process for tissue repair and wound healing (Troncone et al. 2018). Furthermore, TGF-β1 promotes differentiation of M2 macrophages, an anti-inflammatory subset which actively participates in tissue repair and homeostasis and attenuates the macrophage inflammatory response to bacterial products (Troncone et al. 2018). The disruption of the monocyte/macrophage phenotype and a significant shift toward pro-inflammatory polarization of macrophages has been recently reported to be associated with the pathogenesis of periodontal disease (Almubarak et al. 2020). The current data further point to the importance of epithelial TGF-β signaling in periodontitis.

RORA is another computationally predicted differentially expressed gene in gingival epithelium, the lower expression of which in states of health was also validated in epithelial cells isolated from

unrelated gingival tissue samples. Earlier mechanistic studies in human monocytes showed that deletion of RORA leads to activation of the nuclear factor κB (NF-κB) signaling pathway and to downstream induction of pro-inflammatory cytokines such as TNF, IL-1β and IL-6, at both the transcriptional and the protein level (Nejati Moharrami et al. 2018). In the same study, RORA knockout cells were found to produce high levels of pro-IL-1β, even in the absence of lipopolysaccharide challenge. Corroborating these observations, studies using an intestinal epithelium-specific, RORA-deficient mouse model showed that RORA is crucial for maintaining intestinal homeostasis by attenuating NF-κB transcriptional activity and preventing inflammation (Oh et al. 2019). In an earlier study using reverse engineering approaches, we identified RORA as a master regulator of the transcriptional landscape in periodontitis (Sawle et al. 2016). Consistent with these observations, our finding of higher expression of RORA in epithelial cells from healthy gingiva highlights its importance as a potential molecular target for the restoration of epithelial homeostasis and attenuation of innate immunity in periodontitis.

Pathway enrichment analysis of genes predicted by PSEA to be differentially expressed in gingival epithelium showed enrichment of TH17 signaling, AGE-RAGE receptor signaling, and the epithelial-mesenchymal transition (EMT) signaling pathways. Earlier work has demonstrated the involvement of TH17 cells and their signature cytokine profiles in the pathogenesis of periodontitis (Gaffen and Hajishengallis 2008), whereas AGE-RAGE signaling plays a pivotal role in the pathogenesis of both periodontitis and diabetes mellitus (Lalla et al. 2000; Lalla et al. 2001). Higher expression of the receptor for AGEs, RAGE, has been reported in periodontitis-affected gingival tissues, and in the peripheral blood of patients with periodontitis when compared to periodontally healthy individuals; circulating soluble forms of RAGE were proposed as

biomarkers for the presence and severity/extent of periodontitis (Detzen et al. 2019). Challenge of epithelial cells with AGEs was found to result in the phosphorylation of ERK, p38, and subsequent activation of NF-κB (Kido et al. 2020). The third significantly enriched pathway in epithelial cells (EMT) represents a cellular process characterized by changes in transcriptional and proteomic changes that result in the trans-differentiation of the epithelial phenotype to a mesenchymal phenotype. Indeed, PSEA-predicted differential expression of MAPK13, OCL, TGFB and COL4A2 in epithelial cells, all of which are associated with EMT (Scanlon et al. 2013). These findings point to a possible mechanistic overlap between the transcriptional landscape of periodontitis and carcinogenesis which is intriguing and warrants further investigation.

We also predicted and independently validated the downregulation of CERS3 in B cells isolated from periodontitis-affected tissue compared to healthy gingiva. CERS3 is a member of ceramide synthetases protein family, and ceramide is an important signaling molecule in sphingolipid metabolism (Levy and Futerman 2010). Ceramides are present in the cytoplasm of host cells and play essential roles in orchestrating immune responses (Albeituni and Stiban 2019). Recently, a diminished expression of acid ceramidase in periodontal lesions as well as in *Porhyromonas gingivalis*-stimulated epithelial cells *in vitro* was reported (Azuma et al. 2018). Furthermore, overexpression of acid ceramidase in epithelial cells resulted in attenuation of pro-inflammatory immune response and apoptosis in response to challenge by *P. gingivalis*, highlighting a possible anti-inflammatory role of ceramides in gingival tissue (Azuma et al. 2018).

Pathway analysis on predicted differentially expressed genes in fibroblasts showed enrichment of VEGFA-VEGFR2 and Senescence and Autophagy pathways. Activation of VEGF/VEGFR2

axis has been reported in periodontal disease, and high angiogenesis activity in periodontal lesions was correlated with VEGF expression in the stroma (Vladau et al. 2016). Similar to other chronic inflammatory diseases, periodontitis has been associated with autophagic alterations (Zhuang et al. 2016). Increased levels of autophagy gene expression and high levels of mitochondrial reactive oxygen species production in peripheral blood mononuclear cells were observed in patients with periodontitis (Bullon et al. 2012). We also found that the Sterol Regulatory Element-Binding Protein Signaling was among the dysregulated pathways in endothelial cells. SREBP1C is a key lipogenic transcription factor which regulates cholesterol and fatty acid metabolism and synthesis (Wang et al. 2015). Activation and higher levels of SREBP1C has been reported in periodontal disease-affected tissue in patients with diabetes (Kuo et al. 2016). Upregulation of SREBP1C was critical for induction of NLRP3, an inflammasome component, by high-glucose-treated *P. gingivalis* (Kuo et al. 2016). Convergence of these important pathways and their biological relevance to periodontal disease warrant further investigation.

Collectively, our results demonstrate the robustness of the PSEA in the decomposition of gingival tissue transcriptomes, and its ability to identify differentially regulated transcripts in particular cellular constituents. These genes may serve as candidates for further investigation with respect to their roles in the pathogenesis of periodontitis.

AUTHORS CONTRIBUTIONS

FMH contributed to design, experimental data acquisition and interpretation, drafted and critically revised the manuscript; RAF contributed to design, bioinformatics data analysis and interpretation, and critically revised the manuscript; SA contributed to experimental data acquisition; AS contributed to bioinformatics data analysis; MK contributed to data acquisition; AK contributed to design, bioinformatics data analysis and interpretation, and critically revised the manuscript; PNP obtained funding for the study and contributed to conception, design, data acquisition and interpretation, drafted and critically revised the manuscript.

CONFLICTS OF ONTEREST

The authors declare no conflicts of interest.

ACKNOWLEDGEMENTS

This work was supported by grants from NIH/NIDCR (DE015649, DE021820 and DE024735) and by an unrestricted gift from Colgate-Palmolive Inc. to PNP, as well as by the National Center for Advancing Translational Sciences (TR000040). The authors would like to thank Dr. Michael J. Zilliox for helpful correspondence regarding the Barcode.

FIGURE LEGENDS

Figure 1. KEGG and WP pathway analysis in epithelial cells.

Significantly enriched pathways (FDR of ≤0.1) based on ≥ 2 PSEA-decomposed differentially expressed genes are presented, along with the adjusted p value, the -log10 of the adjusted p value, and the involved genes.

Figure 2. Validation of differential expression of PSEA-predicted genes in isolated epithelial cells from gingival tissue.

Relative expression levels (ΔCt values) of TGFβ1 assessed through qRT-PCR in matched (n=7 pairs connected by horizontal lines; panel A) and non-matched analyses (n=18; panel B). Relative expression levels of RORA assessed through qRT-PCR in matched (n=8 pairs by horizontal lines; panel C) and non-matched analyses (n=22; panel D). Data are presented as mean and standard error of mean; 18s was used as normalizer. P-values are derived by one-tailed t-tests, for paired (panels A and C) and non-paired observations (panels B and D). Note that higher ΔCt values indicate lower expression.

Figure 3. Validation of differential expression of PSEA-predicted genes in isolated B cells from gingival tissue.

A) Relative expression levels of CERS3 (ΔCt values) assessed through qRT-PCR (n=14). B) Relative expression levels of CAMSAP1 (ΔCt values) assessed through qRT-PCR (n=12). Data are presented as mean and standard error of mean. 18s was used as normalizer. P-values are derived using one-tailed t-tests for non-paired observations. Note that higher ΔCt values indicate lower expression.

REFERENCES

Akaike H. 1974. Information theory and an extension of the maximum likelihood principle. Second International Symposium on Information Theory. 1(1):267-281.

Albeituni S, Stiban J. 2019. Roles of ceramides and other sphingolipids in immune cell function and inflammation. Adv Exp Med Biol. 1161:169-191.

Almubarak A, Tanagala KKK, Papapanou PN, Lalla E, Momen-Heravi F. 2020. Disruption of monocyte and macrophage homeostasis in periodontitis. Front Immunol. 11:330.

Azuma MM, Balani P, Boisvert H, Gil M, Egashira K, Yamaguchi T, Hasturk H, Duncan M, Kawai T, Movila A. 2018. Endogenous acid ceramidase protects epithelial cells from porphyromonas gingivalis-induced inflammation in vitro. Biochem Biophys Res Commun. 495(4):2383-2389.

Bullon P, Cordero MD, Quiles JL, Ramirez-Tortosa Mdel C, Gonzalez-Alonso A, Alfonsi S, Garcia-Marin R, de Miguel M, Battino M. 2012. Autophagy in periodontitis patients and gingival fibroblasts: Unraveling the link between chronic diseases and inflammation. BMC Med. 10:122.

Cekici A, Kantarci A, Hasturk H, Van Dyke TE. 2014. Inflammatory and immune pathways in the pathogenesis of periodontal disease. Periodontol 2000. 64(1):57-80.

Demmer RT, Behle JH, Wolf DL, Handfield M, Kebschull M, Celenti R, Pavlidis P, Papapanou PN. 2008. Transcriptomes in healthy and diseased gingival tissues. J Periodontol. 79(11):2112-2124.

Deng Q, Ramskold D, Reinius B, Sandberg R. 2014. Single-cell rna-seq reveals dynamic, random monoallelic gene expression in mammalian cells. Science. 343(6167):193-196.

Denney L, Byrne AJ, Shea TJ, Buckley JS, Pease JE, Herledan GM, Walker SA, Gregory LG, Lloyd CM. 2015. Pulmonary epithelial cell-derived cytokine tgf-beta1 is a critical cofactor for enhanced innate lymphoid cell function. Immunity. 43(5):945-958.

Detzen L, Cheng B, Chen CY, Papapanou PN, Lalla E. 2019. Soluble forms of the receptor for advanced glycation endproducts (rage) in periodontitis. Scientific reports. 9(1):8170.

Ebersole JL, Dawson DR, 3rd, Morford LA, Peyyala R, Miller CS, Gonzalez OA. 2013. Periodontal disease immunology: 'Double indemnity' in protecting the host. Periodontol 2000. 62(1):163-202.

Esumi S, Wu SX, Yanagawa Y, Obata K, Sugimoto Y, Tamamaki N. 2008. Method for single-cell microarray analysis and application to gene-expression profiling of gabaergic neuron progenitors. Neurosci Res. 60(4):439-451.

Gaffen SL, Hajishengallis G. 2008. A new inflammatory cytokine on the block: Re-thinking periodontal disease and the th1/th2 paradigm in the context of th17 cells and il-17. J Dent Res. 87(9):817-828.

Guillaumet-Adkins A, Rodriguez-Esteban G, Mereu E, Mendez-Lago M, Jaitin DA, Villanueva A, Vidal A, Martinez-Marti A, Felip E, Vivancos A et al. 2017. Single-cell transcriptome conservation in cryopreserved cells and tissues. Genome Biol. 18(1):45.

Heath JR, Ribas A, Mischel PS. 2016. Single-cell analysis tools for drug discovery and development. Nat Rev Drug Discov. 15(3):204-216.

Horie M, Yamaguchi Y, Saito A, Nagase T, Lizio M, Itoh M, Kawaji H, Lassmann T, Carninci P, Forrest AR et al. 2016. Transcriptome analysis of periodontitis-associated fibroblasts

by cage sequencing identified dlx5 and runx2 long variant as novel regulators involved in periodontitis. Scientific reports. 6:33666.

Howe KL, Reardon C, Wang A, Nazli A, McKay DM. 2005. Transforming growth factor-beta regulation of epithelial tight junction proteins enhances barrier function and blocks enterohemorrhagic escherichia coli o157:H7-induced increased permeability. Am J Pathol. 167(6):1587-1597.

Hunt-Newbury R, Viveiros R, Johnsen R, Mah A, Anastas D, Fang L, Halfnight E, Lee D, Lin J, Lorch A et al. 2007. High-throughput in vivo analysis of gene expression in caenorhabditis elegans. PLoS biology. 5(9):e237.

Kebschull M, Guarnieri P, Demmer RT, Boulesteix AL, Pavlidis P, Papapanou PN. 2013. Molecular differences between chronic and aggressive periodontitis. J Dent Res. 92(12):1081-1088.

Kido R, Hiroshima Y, Kido JI, Ikuta T, Sakamoto E, Inagaki Y, Naruishi K, Yumoto H. 2020. Advanced glycation end-products increase lipocalin 2 expression in human oral epithelial cells. J Periodontal Res.

Kinane DF, Stathopoulou PG, Papapanou PN. 2017. Periodontal diseases. Nat Rev Dis Primers. 3:17038.

Kuhn A, Thu D, Waldvogel HJ, Faull RL, Luthi-Carter R. 2011. Population-specific expression analysis (psea) reveals molecular changes in diseased brain. Nat Methods. 8(11):945-947.

Kuo HC, Chang LC, Chen TC, Lee KC, Lee KF, Chen CN, Yu HR. 2016. Sterol regulatory element-binding protein-1c regulates inflammasome activation in gingival fibroblasts infected with high-glucose-treated porphyromonas gingivalis. Front Cell Infect Microbiol. 6:195.

Lalla E, Lamster IB, Feit M, Huang L, Spessot A, Qu W, Kislinger T, Lu Y, Stern DM, Schmidt AM. 2000. Blockade of rage suppresses periodontitis-associated bone loss in diabetic mice. J Clin Invest. 105(8):1117-1124.

Lalla E, Lamster IB, Stern DM, Schmidt AM. 2001. Receptor for advanced glycation end products, inflammation, and accelerated periodontal disease in diabetes: Mechanisms and insights into therapeutic modalities. Ann Periodontol. 6(1):113-118.

Levy M, Futerman AH. 2010. Mammalian ceramide synthases. IUBMB life. 62(5):347-356.

McCall MN, Jaffee HA, Zelisko SJ, Sinha N, Hooiveld G, Irizarry RA, Zilliox MJ. 2014. The gene expression barcode 3.0: Improved data processing and mining tools. Nucleic Acids Res. 42(1):D938-943.

Nejati Moharrami N, Bjorkoy Tande E, Ryan L, Espevik T, Boyartchuk V. 2018. Roralpha controls inflammatory state of human macrophages. PLoS ONE. 13(11):e0207374.

Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M, Alizadeh AA. 2015. Robust enumeration of cell subsets from tissue expression profiles. Nat Methods. 12(5):453-457.

Oh SK, Kim D, Kim K, Boo K, Yu YS, Kim IS, Jeon Y, Im SK, Lee SH, Lee JM et al. 2019. Roralpha is crucial for attenuated inflammatory response to maintain intestinal homeostasis. Proc Natl Acad Sci U S A. 116(42):21140-21149.

Papapanou PN, Behle JH, Kebschull M, Celenti R, Wolf DL, Handfield M, Pavlidis P, Demmer RT. 2009. Subgingival bacterial colonization profiles correlate with gingival tissue gene expression. BMC Microbiol. 9:221.

Reimand J, Kull M, Peterson H, Hansen J, Vilo J. 2007. G:Profiler--a web-based toolset for functional profiling of gene lists from large-scale experiments. Nucleic Acids Res. 35(Web Server issue):W193-200.

Sawle AD, Kebschull M, Demmer RT, Papapanou PN. 2016. Identification of master regulator genes in human periodontitis. J Dent Res. 95(9):1010-1017.

Scanlon CS, Van Tubergen EA, Inglehart RC, D'Silva NJ. 2013. Biomarkers of epithelial-mesenchymal transition in squamous cell carcinoma. J Dent Res. 92(2):114-121.

Takayanagi H. 2005. Inflammatory bone destruction and osteoimmunology. J Periodontal Res. 40(4):287-293.

Troncone E, Marafini I, Stolfi C, Monteleone G. 2018. Transforming growth factor-beta1/smad7 in intestinal immunity, inflammation, and cancer. Front Immunol. 9:1407.

Wang Y, Viscarra J, Kim SJ, Sul HS. 2015. Transcriptional regulation of hepatic lipogenesis. Nat Rev Mol Cell Biol. 16(11):678-689.

Vladau M, Cimpean AM, Balica RA, Jitariu AA, Popovici RA, Raica M. 2016. Vegf/vegfr2 axis in periodontal disease progression and angiogenesis: Basic approach for a new therapeutic strategy. In Vivo. 30(1):53-60.

Zhuang H, Ali K, Ardu S, Tredwin C, Hu B. 2016. Autophagy in dental tissues: A double-edged sword. Cell Death Dis. 7:e2192.

Table 1. PSEA-decomposed, cell type-specific gene expression profiles that fulfilled all filtering steps*.

Probe ID	Gene symbol	Log2FC	Ref p value	Diff p value
Epithelial cells				
210479_s_at	RORA	-0.67	1.1E-10	1.1E-07
205637_s_at	SH3GL3	-0.43	2.1E-21	4.5E-07
227309_at	YOD1	-0.45	3.5E-21	1.0E-06
211966_at	COL4A2	-1.59	8.7E-04	3.2E-06
223895_s_at	EPN3	-0.42	1.4E-13	3.3E-05
222190_s_at	C16orf58	-1.46	3.4E-02	1.9E-03
225510_at	OAF	-0.57	2.2E-05	2.7E-03
226632_at	CYGB	-1.26	3.9E-02	5.7E-03
203085_s_at	TGFB1	-0.69	2.1E-03	0.01
209216_at	WDR45	-0.46	1.0E-04	0.01
Fibroblasts				
208872_s_at	REEP5	0.95	0.047	1.10E-03
219315_s_at	TMEM204	-0.71	1.60E-04	6.80E-03
202828_s_at	MMP14	0.57	2.50E-03	0.012
208851_s_at	THY1	-0.46	3.10E-09	3.50E-03
Endothelial cells				
225369_at	ESAM	-0.41	4.5E-26	5.7E-07
228339_at	ECSCR	-0.44	5.5E-24	3.2E-06
215535_s_at	AGPAT1	-0.65	8.7E-08	1.0E-03
212494_at	TNS2	-0.52	2.1E-07	6.0E-03
209166_s_at	MAN2B1	-0.68	9.73E-05	0.01
Neutrophils				
226907_at	PPP1R14C	-1.14	5.7E-09	2.4E-10
223694_at	TRIM7	-0.99	1.1E-07	1.9E-08
202428_x_at	DBI	-3.19	1.3E-02	1.3E-05

232116_at	GRHL3	-0.74	2.7E-07	2.5E-05
227736_at	C10orf99	-1.15	2.1E-04	4.5E-05
220013_at	EPHX3	-0.95	3.2E-05	5.4E-05
204203_at	CEBPG	-0.39	5.2E-14	6.3E-05
230769_at	DENND2C	-0.97	7.9E-05	1.1E-04
214626_s_at	GANAB	0.83	1.0E-02	1.1E-04
228587_at	FAM83G	-0.43	4.8E-12	2.2E-04
204616_at	UCHL3	-0.61	3.1E-07	3.1E-04
209311_at	BCL2L2	-1.2	3.1E-03	4.7E-04
201315_x_at	IFITM2	0.75	2.0E-02	1.2E-03
224615_x_at	HM13	0.55	1.0E-02	0.02
Plasma cells				
212890_at	SLC38A10	-1.17	1.3E-11	1.4E-05
55093_at	CHPF2	-1.3	1.4E-09	4.2E-05
206593_s_at	MED22	-2.75	1.5E-03	2.6E-03
204158_s_at	TCIRG1	-1.21	2.5E-04	0.02
200644_at	MARCKSL1	-1.49	3.9E-03	0.04
202369_s_at	TRAM2	-0.55	3.7E-08	0.05
B cells				
202539_s_at	HMGCR	-0.89	1.20E-09	4.00E-05
204552_at	INPP4A	0.86	0.01	3.00E-03
210785_s_at	THEMIS2	0.85	0.02	6.00E-03
204912_at	IL10RA	0.6	2.20E-04	6.00E-03
212712_at	CAMSAP1	-0.89	9.30E-05	8.00E-03
220306_at	FAM46C	0.83	0.02	0.01
1554252_a_at	CERS3	-1.65	0.02	0.02
206896_s_at	GNG7	0.78	0.04	0.04

* p-value of presence of the gene in the cell type that is differentially expressed <0.05; p-value of differential expression in periodontitis-affected versus healthy gingiva <0.05; |log2FC differential expression| >0.4, and confidence coefficient (CC)=1.

Log2FC: Log2-based fold change of expression in periodontitis affected- over healthy gingival tissues

Ref p value: p-value for expression of the particular probe at the specific cell type

Diff p value: p-value for the differential expression of the particular probe between periodontitis-affected and healthy gingival tissues at the specific cell type

Table 2. Over-represented KEGG and WP pathways based on ≥ 2 PSEA-decomposed differentially expressed genes between periodontitis-affected and healthy gingival tissues, by cell type

Pathway	Pathway ID	Adjusted p Value	Genes
Epithelial cells			
Th17 cell differentiation	KEGG:04659	0.03031515	RORA, MAPK13, TGFB1
AGE-RAGE signaling pathway in diabetic complications	KEGG:04933	0.03031515	COL4A2, MAPK13, TGFB1
Relaxin signaling pathway	KEGG:04926	0.03701637	COL4A2, MAPK13, TGFB1
Pathogenic Escherichia coli infection	KEGG:05130	0.09286103	OCLN, MAPK13, TUBB2A
Inflammatory bowel disease (IBD)	KEGG:05321	0.09286103	RORA, TGFB1
Epithelial to mesenchymal transition in colorectal cancer	WP:WP4239	0.00876677	COL4A2, OCLN, MAPK13, TGFB1
Cannabinoid receptor signaling	WP:WP3869	0.04737078	MAPK13, CYP2C9
Hepatitis C and Hepatocellular Carcinoma	WP:WP3646	0.06769704	COL4A2, TGFB1
Pathogenic Escherichia coli infection	WP:WP2272	0.06769704	OCLN, TUBB2A
Fibroblasts			
VEGFA-VEGFR2 Signaling	WP:WP3888	0.04963475	CALU, MMP14, PBXIP1
Senescence and Autophagy in Cancer	WP:WP615	0.04963475	IGFBP5, MMP14
Endothelial cells			
Fat digestion and absorption	KEGG:04975	0.05687131	AGPAT1, SCARB1
Sterol Regulatory Element-Binding Proteins (SREBP) signaling	WP:WP1982	0.060402348	SEC24A, SCARB1
Neutrophils			
Human Complement System	WP:WP2806	0.080752415	SELPLG, LAMC1
Nuclear Receptors Meta-Pathway	WP:WP2882	0.080752415	PPP1R14C, DBI, CYP2C9
Plasma cells			
Protein processing in endoplasmic reticulum	KEGG:04141	0.005626135	WFS1, RRBP1, PREB
B cells			
Human cytomegalovirus infection	KEGG:05163	0.071917892	IL10RA, GNG7, TAPBP

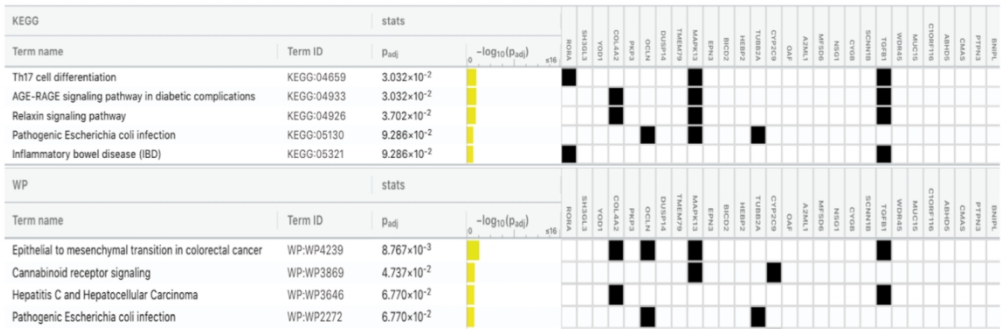

Figure 1. KEGG and WP pathway analysis in epithelial cells.
Significantly enriched pathways (FDR of ≤0.1) based on ≥ 2 PSEA-decomposed differentially expressed genes are presented, along with the adjusted p value, the -log10 of the adjusted p value, and the involved genes.

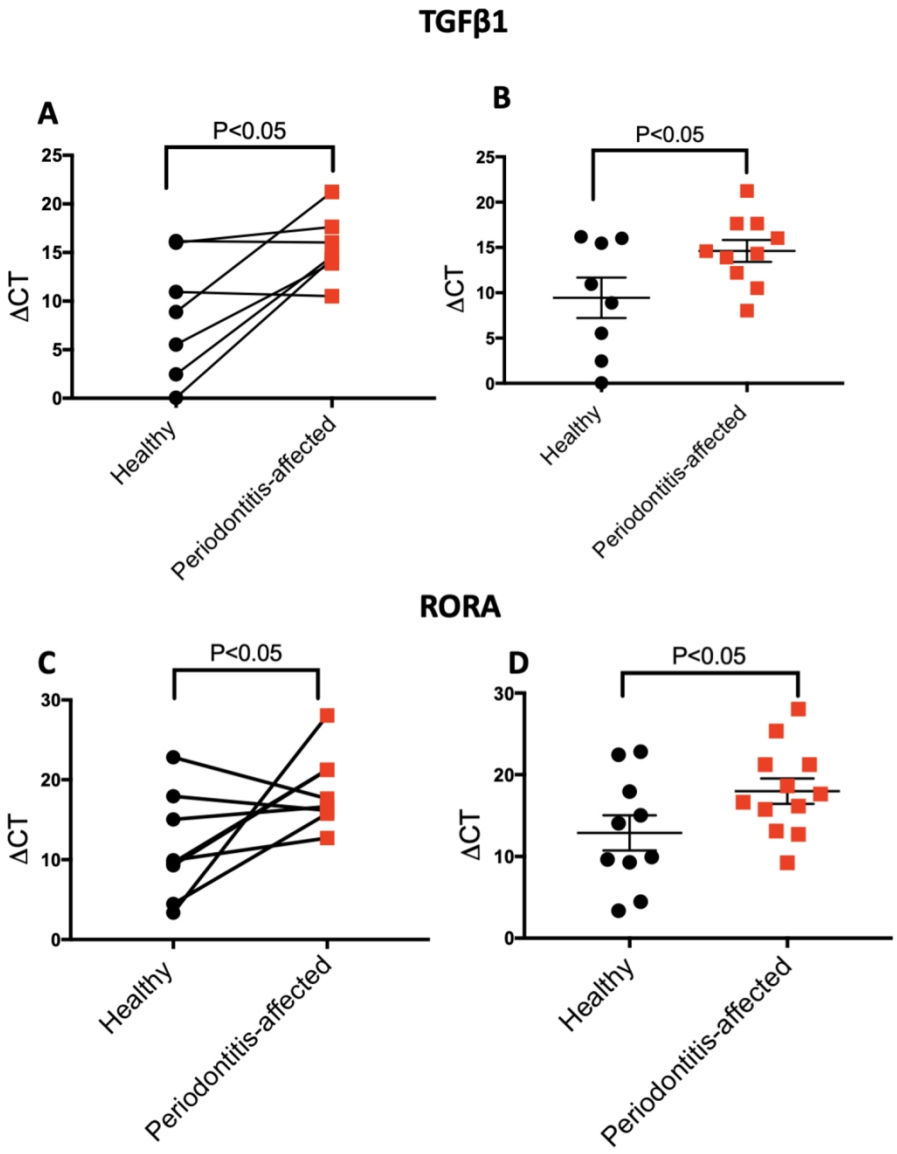

Figure 2. Validation of differential expression of PSEA-predicted genes in isolated epithelial cells from gingival tissue.
Relative expression levels (□Ct values) of TGF□1 assessed through qRT-PCR in matched (n=7 pairs connected by horizontal lines; panel A) and non-matched analyses (n=18; panel B). Relative expression levels of RORA assessed through qRT-PCR in matched (n=8 pairs by horizontal lines; panel C) and non-matched analyses (n=22; panel D). Data are presented as mean and standard error of mean; 18s was used as normalizer. P-values are derived by one-tailed t-tests, for paired (panels A and C) and non-paired observations (panels B and D). Note that higher □Ct values indicate lower expression.

136x172mm (300 x 300 DPI)

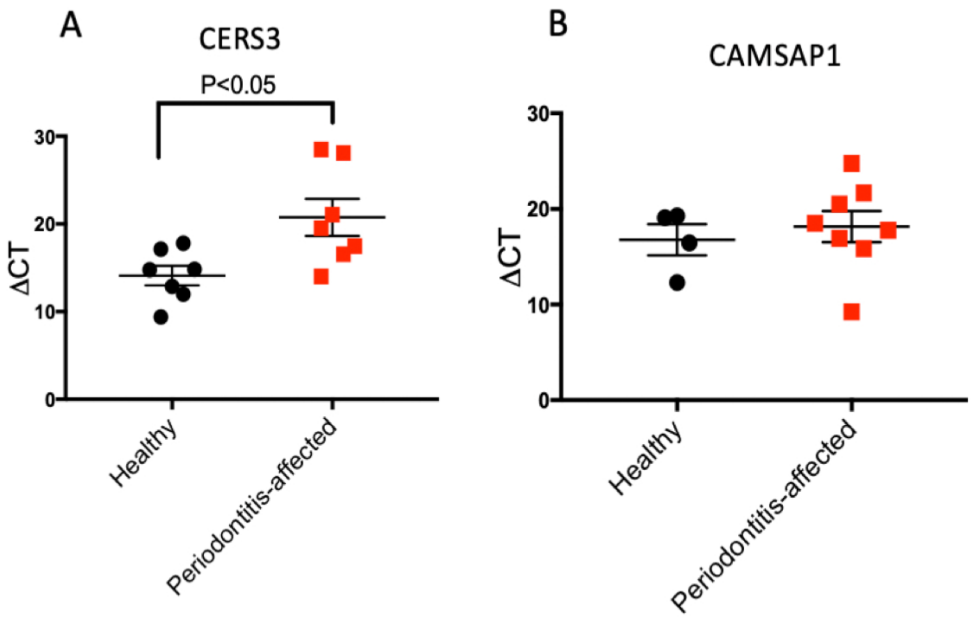

Figure 3. Validation of differential expression of PSEA-predicted genes in isolated B cells from gingival tissue. A) Relative expression levels of CERS3 (□Ct values) assessed through qRT-PCR (n=14). B) Relative expression levels of CAMSAP1 (□Ct values) assessed through qRT-PCR (n=12). Data are presented as mean and standard error of mean. 18s was used as normalizer. P-values are derived using one-tailed t-tests for non-paired observations. Note that higher □Ct values indicate lower expression.

SUPPLEMENTAL METHODS

PSEA analysis

Gene expression data set

The dataset included whole tissue transcriptomes from a total of 310 gingival tissue samples obtained from 120 patients with periodontitis, comprising interproximal papillae that were either periodontitis-affected [241 samples showing bleeding on probing (BoP), probing depth (PD) ≥4mm and clinical attachment level (CAL≥3mm)] or clinically healthy (69 samples; with no BoP, PD≤4 mm, and CAL≤2 mm). Inclusion criteria, demographics, sample characteristics and processing pipeline have been published previously (Kebschull et al. 2013).

Five outlier arrays (all from periodontitis-affected samples) were identified by the GNUSE method and removed, as earlier described (McCall et al. 2014). Data was first normalized using GCRMA (Wu et al. 2004; Wu and Irizarry 2005) and further between batches using COMBAT (Johnson et al. 2007) implemented in the SVA R package, as earlier described (Sawle et al. 2016). Whenever two periodontitis-affected samples were available from the same donor, their intensities were averaged. Thus, the dataset further analyzed comprised 118 periodontitis-associated and 69 healthy gingival tissue samples.

Summary of the Barcode Methodology

The Gene Expression Barcode 3.0 method (McCall et al. 2011; McCall et al. 2014) takes as its input the expression of each probeset for a given chip platform from samples from a wide variety of cell and tissue types, normalized by the frozen Robust Multichip Algorithm

(fRMA; McCall et al. 2010). The method infers an expression threshold for each chip by an hierarchical (Gelman and Hill 2006) Bayesian (Gelman et al. 2014) mixture model (Everitt and Hand 1981) which is an adaptation of the Probability of Expression model (POE; Parmigiani et al. 2002). The probability of expression of a given probeset in each cell or tissue type is then taken to be the fraction of samples of that cell or tissue type whose expression is greater than or equal to the expression threshold of that probeset.

Filtering of marker probesets (genes)

The list of marker probesets were filtered further based upon gene expression data as follows: (i) if two probesets representing the same gene had a Pearson correlation coefficient of<0.7, one of them was discarded; (ii) if two probesets which were candidate markers for different cell types had a Pearson correlation coefficient p-value<0.05, one was eliminated (this criterion guaranteed that the markers for different cell types were not correlated in our models); (iii) the variance inflation factor (VIF) described below for all of the probesets within a marker for a cell type was<10; (iv) probesets were further discarded based upon known expression in other cellular subtypes in the gingiva.

Initial filtering of probesets in the differential expression analysis

Probesets were removed from the differential expression analyses based upon the following criteria listed in the PSEA publication (Kuhn et al. 2011) : (i) probesets used as markers, to avoid circular reasoning; (ii) probesets with an adjusted coefficient of determination

R^2<0.6 for the best model, indicating a poor fit; and (iii) probesets whose intercepts were>0.5 their average expression, indicating that their expression did not vary with the marker for a cell type.

Cell type-specific filtering of probesets in the differential expression analysis

Additional filtering criteria were applied to select probesets of interest: (i) p-value of presence of the gene in the cell type that is differentially expressed<0.05; (ii) p-value of differential expression in periodontitis-affected versus healthy gingiva<0.05; (iii) absolute log2FC differential expression of>0.4, and (iv) confidence coefficient (CC); i.e., the fraction of models with the same cell-type differentially expressed within 2 AIC of the best model, =1. Additional filtered-out probesets included those with negative coefficients of presence (corresponding to negative concentration) and those whose differential expression led to net negative concentration. Redundant probesets for the same gene were removed.

Validation of PSEA predicted genes

Gingival tissue harvesting and preparation of single cell suspensions

Gingival tissue samples were harvested from patients in conjunction with periodontal surgical procedures (pocket elimination/reduction surgery, crown lengthening or tooth extraction) after approval by the Columbia University Medical Center Institutional Review Board (Protocol # AAAR0526). Patients were recruited at the Clinic of Graduate Periodontics of the College of Dental Medicine and informed consent was obtained. All patients were systemically healthy, non-pregnant, non-smokers who had not used antibiotics or anti-

inflammatory drugs for the preceding 3-month period, as in our previous publications (Kebschull et al. 2013). Interproximal papillae included in the surgical area were harvested and originated either from areas affected by periodontitis, i.e., an interproximal site with PD ≥ 5 mm, with concomitant CAL ≥ 3 mm, presence of radiographic bone, and BoP or from clinically healthy sites, (i.e., sites with PD≤ 3 mm, no CAL, no radiographic bone loss and no BoP). From each patient one periodontitis-affected gingival tissue sample and one healthy tissue sample were harvested (15 pairs, n=30). Gingival tissue samples were processed to form a single-cell suspension using an established laboratory protocol (Almubarak et al. 2020), using a commercially available tissue dissociation kit (Miltenyi Biotech, USA). Samples were kept in Dulbecco's Phosphate-Buffered Saline, and minced into small pieces after washing with saline. Samples were processed in c-tubes (Miltenyi Biotech, USA) which contained 2.35 ml of RPMI 1640 (Thermo Fisher, USA) and 100 µl of Enzyme H, 50 µl of Enzyme R, and 12.5 µl of Enzyme A (Miltenyi Biotech, USA), as recommended by the manufacturer. A gentle MACS Dissociator (Miltenyi Biotech, USA) was used for tissue disruption and enzymatic digestion at 37°C as recommended by the manufacturer. The combination of this mechanical and enzymatic digestion leads to formation single cell suspensions, while maintaining cellular integrity. Cell suspensions were filtered with 70-micrometer filters and each sample was washed with 15 ml of RPMI 1640 solution. Cells were pelleted by centrifugation for 7 minutes (relative centrifugal field: 0.4) and cryopreserved immediately. The pellet was mixed with a freezing medium, which contained 90% of fetal bovine serum (FBS; Corning, USA) and 10% of dimethyl sulfoxide (DMSO; Sigma-Aldrich, USA). The samples were transferred to isopropanol chambers in a -80°C freezer and then transferred into liquid nitrogen within 24h. This protocol has been shown to have minimal effects on the transcriptional profiles after cell revival (Guillaumet-Adkins et al. 2017).

Immuno-magnetic separation of epithelial cells and B cells

All cryopreserved samples were revived from liquid nitrogen and viability was assessed using Trypan blue staining using a TC20 Automated Cell Counter. Approximately 5-6 x 10^8 cells were aliquoted in 300 µl of autoMACS® Running Buffer (Miltenyi Biotec, USA) and 100 µl of FcR blocking reagent (Miltenyi Biotec, USA) was added and shaken gently for 5 minutes. 100 µl of CD326 (EpCAM) magnetic MicroBeads for isolation of epithelial cells or CD19 magnetic MicroBeads (Miltenyi Biotec, USA) for isolation of B cells were added to the suspension and incubated on a rocking platform for 30 minutes at 4°C. EpCAM is an established marker for epithelial cell isolation and has been reported to be highly expressed in gingival epithelial cells (Hasegawa et al. 2017; Huang et al. 2018; Balfe et al. 2018; Hyun et al. 2019). The cells were washed by adding 5 ml of buffer and centrifuged at 300×g for 10 minutes. The supernatants were aspirated and the cells were suspended in 500 µl of buffer. An LS column (Miltenyi Biotec, USA) was placed in the magnetic field and prepared by rinsing with 500µl of buffer. The cell suspension was applied gently from the sides onto the column and was washed three times using 500 µl of buffer. The column was removed from the separator and placed on a collection tube provided by the kit. 1ml of buffer was pipetted onto the column and the magnetically labeled cells were pushed by the plunger and isolated in a new tube. The isolated cells were pelleted by centrifugation at 800xg for 7 minutes, 500 µl TRI reagent (Zymo Research, USA) were added to the cell pellet and mixed well by pipetting, and the samples were kept at -80°C for RNA isolation.

RNA isolation

The Direct-zol RNA Microprep Kit (Zymo Research, USA) was used for RNA isolation according to the manufacturer's instructions. RNA was eluted in 30 µl RNase-free water. The quantity and quality of the RNA (260/280 and 260/230 ratios) were determined using a NanoDrop 1000 device.

Real-time quantitative reverse transcription PCR (RT-qPCR)

Prior to selection of probesets considered for PCR validation, Component and Residual (COR) plots, i.e., plots of the expression of a probeset predicted by the model against the expression of the marker genes with the error added were generated and the linearity of the plots was examined. Variance Inflation Factors (VIFs), that estimate the effect of collinearity on the final fit, were also considered and were required to be <10 for probesets further considered for PCR validation (Fox 2008; Fox and Weisberg 2011). We finally selected for validation two genes predicted by PSEA as differentially expressed in epithelial cells (RORA and TGF-β1) and two in B cells (CERS3 and CAMSAP1). cDNA was transcribed from 75 ng of total RNA utilizing a SuperScript VILO cDNA Synthesis Kit (Invitrogen, USA) in a final volume of 20 µl. The cycle for cDNA synthesis was as follows: 10 min, 25°C; 120 min, 37°C; 5 min 85°C. SYBR-Green-based real-time quantitative PCR (RT-qPCR) was performed using the SYBR™ Select Master Mix (Applied Biosystems, USA) and the CDX96 Real Time PCR Detection System, following a standardized protocol. The thermal cycling conditions were as follows: 2 min, 50°C; 2 min, 95°C; 15 sec 95°C; and 1 min 60°C. Primers were designed for each of the four genes using Primer-BLAST, and

are presented in Supplementary Table 2. mRNA levels were normalized against 18s (internal control) and relative levels were calculated comparing the ΔCt values. One-tailed t-tests, for paired or unpaired observations, as appropriate, based on the availability of pairs of periodontitis-affected/healthy gingival tissue samples with good RNA quality from the same donor, were carried out to test differential expression between gingival health and periodontitis. Statistical significance was defined as p value less than 0.05. Data are presented as mean ± standard error of mean (Figures 2 and 3).

REFERENCES

Almubarak A, Tanagala KKK, Papapanou PN, Lalla E, Momen-Heravi F. 2020. Disruption of monocyte and macrophage homeostasis in periodontitis. Front Immunol. 11:330.

Balfe A, Lennon G, Lavelle A, Docherty NG, Coffey JC, Sheahan K, Winter DC, O'Connell PR. 2018. Isolation and gene expression profiling of intestinal epithelial cells: Crypt isolation by calcium chelation from in vivo samples. Clin Exp Gastroenterol. 11:29-37.

Everitt BS, Hand DJ. 1981. Finite mixture distributions. London: Chapman and Hall.

Fox J. 2008. Applied regression analysis and generalized linear models. Thousand Oaks: SAGE Publications, Inc.

Fox J, Weisberg S. 2011. An r companion to applied regression. Sage.

Gelman A, Hill J. 2006. Data analysis using regression and multilevel/hierarchical models. Cambridge, UK: Cambridge University Press.

Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. 2014. Bayesian data analysis. Boca Raton, Florida, USA: CRC Prress.

Guillaumet-Adkins A, Rodriguez-Esteban G, Mereu E, Mendez-Lago M, Jaitin DA, Villanueva A, Vidal A, Martinez-Marti A, Felip E, Vivancos A et al. 2017. Single-cell transcriptome conservation in cryopreserved cells and tissues. Genome Biol. 18(1):45.

Hasegawa K, Sato A, Tanimura K, Uemasu K, Hamakawa Y, Fuseya Y, Sato S, Muro S, Hirai T. 2017. Fraction of mhcii and epcam expression characterizes distal lung epithelial cells for alveolar type 2 cell isolation. Respir Res. 18(1):150.

Huang L, Yang Y, Yang F, Liu S, Zhu Z, Lei Z, Guo J. 2018. Functions of epcam in physiological processes and diseases (review). International journal of molecular medicine. 42(4):1771-1785.

Hyun SY, Mun S, Kang KJ, Lim JC, Kim SY, Han K, Jang YJ. 2019. Amelogenic transcriptome profiling in ameloblast-like cells derived from adult gingival epithelial cells. Scientific reports. 9(1):3736.

Johnson WE, Li C, Rabinovic A. 2007. Adjusting batch effects in microarray expression data using empirical bayes methods. Biostatistics. 8(1):118-127.

Kebschull M, Guarnieri P, Demmer RT, Boulesteix AL, Pavlidis P, Papapanou PN. 2013. Molecular differences between chronic and aggressive periodontitis. J Dent Res. 92(12):1081-1088.

Kuhn A, Thu D, Waldvogel HJ, Faull RL, Luthi-Carter R. 2011. Population-specific expression analysis (psea) reveals molecular changes in diseased brain. Nat Methods. 8(11):945-947.

McCall MN, Bolstad BM, Irizarry RA. 2010. Frozen robust multiarray analysis (frma). Biostatistics. 11(2):242-253.

McCall MN, Uppal K, Jaffee HA, Zilliox MJ, Irizarry RA. 2011. The gene expression barcode: Leveraging public data repositories to begin cataloging the human and murine transcriptomes. Nucleic Acids Res. 39(Database issue):D1011-1015.

McCall MN, Jaffee HA, Zelisko SJ, Sinha N, Hooiveld G, Irizarry RA, Zilliox MJ. 2014. The gene expression barcode 3.0: Improved data processing and mining tools. Nucleic Acids Res. 42(1):D938-943.

Parmigiani G, Garrett ES, Anbazhagan R, Gabrielson E. 2002. A statistical framework for expression-based molecular classification in cancer. J R Statist Soc B. 64, Part 4:717–736.

Sawle AD, Kebschull M, Demmer RT, Papapanou PN. 2016. Identification of master regulator genes in human periodontitis. J Dent Res. 95(9):1010-1017.

Wu Z, Irizarry RA, Gentleman R, Murillo FM, Spencer F. 2004. A model based background adjustment for oligonucleotide expression. J Am Stat Assoc 2004. 99(4): 909–917.

Wu Z, Irizarry RA. 2005. Stochastic models inspired by hybridization theory for short oligonucleotide arrays. J Comput Biol. 12(6):882-893.

Supplementary Table 1. Marker probesets used in the PSEA analysis

Epithelial cells		Fibroblasts		Endothelial cells		Neutrophils	
Probe id	**Symbol**	**Probe id**	**Symbol**	**Probe id**	**Symbol**	**Probe id**	**Symbol**
1564307_a_at	A2ML1	1555229_a_at	C1S	204677_at	CDH5	206209_s_at	CA4
205623_at	ALDH3A1	231766_s_at	COL12A1	222885_at	EMCN	210789_x_at	CEACAM3
220620_at	C1orf42	231879_at	COL12A1	219436_s_at	EMCN	223552_at	LRRC4
220026_at	CLCA4	212489_at	COL5A1	212951_at	GPR116	207890_s_at	MMP25
224329_s_at	CNFN	202765_s_at	FBN1	203934_at	KDR	1553513_at	VNN3
206642_at	DSG1	221447_s_at	GLT8D2	209087_x_at	MCAM		
219995_s_at	FLJ13841	205422_s_at	ITGBL1	228863_at	PCDH17		
214599_at	IVL	204682_at	LTBP2	221529_s_at	PLVAP		
205470_s_at	KLK11	223690_at	LTBP2	209070_s_at	RGS5		
239381_at	KLK7	212246_at	MCFD2	218353_at	RGS5		
205778_at	KLK7	1557938_s_at	PTRF	206211_at	SELE		
206400_at	LGALS7			204468_s_at	TIE1		
206884_s_at	SCEL						
1554921_a_at	SCEL						
211361_s_at	SERPINB13						
205185_at	SPINK5						
205064_at	SPRR1B						
206008_at	TGM1						
230835_at	UNQ467						
226926_at	ZD52F10						

Monocytes/macrophages		Plasma cells		T cells		B cells	
Probe id	**Symbol**	**Probe id**	**Symbol**	**Probe id**	**Symbol**	**Probe id**	**Symbol**
207270_x_at	CD300C	235965_at	DKFZP434B0335	211861_x_at	CD28	1563469_at	ARID5B
204150_at	STAB1	219910_at	HYPE	206980_s_at	FLT3LG	212715_s_at	MICAL3
38487_at	STAB1	240915_at	IGHV1-69				
		231931_at	PRDM15				

Supplementary Table 2. Primers used in the validation experiments

Gene		
18S	*Forward*	5′-GACCTCATCCCACCTCTCAG-3′
	Reverse	5′-CCATCCAATCGGTAGTAGCG-3′
TGF-β	*Forward*	5’-ACGCAGTACAGCAAGGTCC-3’
	Reverse	5’-GACACAGAGATCCGCAGTCC-3’
RORA	*Forward*	5’-TTGCGGGTGTACCTTGATCC-3’
	Reverse	5’- CTGGCTGCCCCTCAACAATA-3’
CERS3	*Forward*	5’- GGAAGCTTGCTGGAGATTTGC-3’
	Reverse	5’- CAGTACTGGGATGGCAGCAG-3’
CAMSAP1	*Forward*	5’-GAATGATGGCTGCAGTTGGC-3’
	Reverse	5’- GTCATGAGGGTGGGGAATGG-3’

Supplementary Table 3. PSEA -predicted differentially expressed genes by cell type

Epithelial cells

Probe id	Symbol	log2FC	Ref_p	Dif_p	R^2	CC
210479_s_at	RORA	-0.67	1.1E-10	1.1E-07	0.63	1
205637_s_at	SH3GL3	-0.43	2.1E-21	4.5E-07	0.66	1
227309_at	YOD1	-0.45	3.5E-21	1.0E-06	0.62	1
211966_at	COL4A2	-1.59	8.7E-04	3.2E-06	0.77	1
209873_s_at	PKP3	-0.25	1.3E-39	3.4E-06	0.71	0.74
209925_at	OCLN	-0.31	6.5E-30	6.0E-06	0.62	1
203367_at	DUSP14	-0.46	4.2E-16	1.2E-05	0.62	0.88
223544_at	TMEM79	-0.17	2.1E-54	2.6E-05	0.80	1
210059_s_at	MAPK13	-0.27	3.2E-29	3.2E-05	0.63	1
223895_s_at	EPN3	-0.42	1.4E-13	3.3E-05	0.61	1
209203_s_at	BICD2	-0.29	3.0E-26	3.4E-05	0.65	0.86
203430_at	HEBP2	-0.2	3.5E-28	5.9E-04	0.62	0.5
209372_x_at	NA	-0.22	2.1E-29	8.6E-04	0.61	1
216661_x_at	CYP2C9	-0.26	1.5E-21	1.0E-03	0.64	0.6
222190_s_at	C16orf58	-1.46	3.4E-02	1.9E-03	0.64	1
225510_at	OAF	-0.57	2.2E-05	2.7E-03	0.64	1
1553505_at	A2ML1	-0.23	8.6E-25	2.8E-03	0.62	1
219858_s_at	MFSD6	-0.22	7.2E-23	3.2E-03	0.62	1
213533_at	NSG1	-0.26	1.7E-13	4.9E-03	0.65	0.47
226632_at	CYGB	-1.26	3.9E-02	5.7E-03	0.64	1
205464_at	SCNN1B	-0.13	2.4E-38	6.7E-03	0.63	0.5
203085_s_at	TGFB1	-0.69	2.1E-03	0.01	0.63	1
209216_at	WDR45	-0.46	1.0E-04	0.01	0.63	1
227241_at	MUC15	-0.14	5.9E-31	0.02	0.73	1
219476_at	C1orf116	-0.18	2.1E-24	0.02	0.62	0.42

218739_at	ABHD5	-0.17	4.9E-23	0.03	0.65	0.89
218111_s_at	CMAS	-0.17	7.1E-21	0.03	0.66	0.36
203997_at	PTPN3	-0.09	5.5E-37	0.04	0.69	1
1553072_at	BNIPL	-0.13	3.5E-26	0.05	0.61	0.9

Fibroblasts

Probe id	Symbol	log2FC	Ref_p	Dif_p	R^2	CC
201539_s_at	FHL1	0.27	3.7E-26	2.6E-05	0.66	1
211958_at	IGFBP5	0.16	9.5E-23	2.3E-02	0.64	1
201718_s_at	EPB41L2	0.17	2.5E-22	1.6E-02	0.67	0.54
214845_s_at	CALU	0.25	4.3E-14	6.7E-03	0.61	0.39
208851_s_at	THY1	-0.46	3.1E-09	3.5E-03	0.68	1
219315_s_at	TMEM204	-0.71	1.6E-04	6.8E-03	0.73	1
205240_at	GPSM2	-0.68	3.8E-04	0.02	0.60	0.82
208829_at	TAPBP	0.46	8.8E-04	0.02	0.71	0.45
202828_s_at	MMP14	0.57	2.5E-03	0.01	0.64	1
212259_s_at	PBXIP1	0.64	0.01	0.01	0.67	0.5
208872_s_at	REEP5	0.95	0.05	0.001	0.60	1
211633_x_at	IGHG1	0.72	0.05	0.02	0.68	0.25

Endothelial cells

Probe id	Symbol	log2FC	Ref_p	Dif_p	R^2	CC
225369_at	ESAM	-0.41	4.5E-26	5.7E-07	0.76	1
228339_at	ECSCR	-0.44	5.5E-24	3.2E-06	0.68	1
215535_s_at	AGPAT1	-0.65	8.7E-08	1.0E-03	0.65	1
206331_at	CALCRL	0.32	4.9E-14	1.0E-03	0.65	0.22
213131_at	OLFM1	-0.3	7.3E-21	2.0E-03	0.65	0.36

212494_at	TNS2	-0.52	2.1E-07	6.0E-03	0.65	1
201389_at	ITGA5	-0.29	3.0E-15	0.01	0.68	1
212902_at	SEC24A	0.75	0.01	0.01	0.63	0.15
209166_s_at	MAN2B1	-0.68	9.73E-05	0.01	0.67	1
205247_at	NOTCH4	-0.3	1.08E-13	0.02	0.63	1
209474_s_at	ENTPD1	0.5	4.12E-04	0.02	0.68	0.71
1552256_a_at	SCARB1	-0.65	9.11E-05	0.02	0.66	0.45
200827_at	PLOD1	-0.35	5.33E-09	0.04	0.71	1

Neutrophils

Probe id	Symbol	log2FC	Ref_p	Dif_p	R^2	CC
226907_at	PPP1R14C	-1.14	5.7E-09	2.4E-10	0.63	1
223694_at	TRIM7	-0.99	1.1E-07	1.9E-08	0.62	1
202428_x_at	DBI	-3.19	1.3E-02	1.3E-05	0.81	1
232116_at	GRHL3	-0.74	2.7E-07	2.5E-05	0.75	1
227736_at	C10orf99	-1.15	2.1E-04	4.5E-05	0.62	1
220013_at	EPHX3	-0.95	3.2E-05	5.4E-05	0.67	1
204203_at	CEBPG	-0.39	5.2E-14	6.3E-05	0.60	1
230769_at	DENND2C	-0.97	7.9E-05	1.1E-04	0.71	1
214626_s_at	GANAB	0.83	1.0E-02	1.1E-04	0.75	1
228587_at	FAM83G	-0.43	4.8E-12	2.2E-04	0.63	1
204616_at	UCHL3	-0.61	3.1E-07	3.1E-04	0.66	1
209311_at	BCL2L2	-1.2	3.1E-03	4.7E-04	0.60	1
216025_x_at	CYP2C9	-0.77	6.5E-05	7.9E-04	0.63	0.58
201315_x_at	IFITM2	0.75	2.0E-02	1.2E-03	0.70	1
209569_x_at	NSG1	-0.71	1.3E-04	2.6E-03	0.66	0.8
221854_at	PKP1	-0.85	4.2E-03	7.6E-03	0.67	0.5
212702_s_at	BICD2	-0.46	2.1E-05	0.01	0.61	0.93

224615_x_at	HM13	0.55	1.0E-02	0.02	0.68	1
209880_s_at	SELPLG	0.28	2.7E-06	0.02	0.65	0.56
218084_x_at	FXYD5	-0.85	2.0E-02	0.04	0.72	0.5
200770_s_at	LAMC1	0.34	2.2E-04	0.04	0.80	0.6

Monocytes/Macrophages

Probe id	Symbol	log2FC	Ref_p	Dif_p	R^2	CC
210657_s_at	SEPT4	0.75	5.20E-03	1.30E-04	0.66	0.67
220532_s_at	TMEM176B	0.35	1.90E-08	1.10E-03	0.72	0.57
202112_at	VWF	0.26	1.40E-08	0.01	0.73	0.5
204503_at	EVPL	-0.32	2.40E-08	0.02	0.60	0.55
211881_x_at	IGLJ3	0.38	5.20E-05	0.02	0.73	0.18

Plasma cells

Probe id	Symbol	log2FC	Ref_p	Dif_p	R^2	CC
212890_at	SLC38A10	-1.17	1.3E-11	1.4E-05	0.76	1
55093_at	CHPF2	-1.3	1.4E-09	4.2E-05	0.70	1
206593_s_at	MED22	-2.75	1.5E-03	2.6E-03	0.61	1
202908_at	WFS1	-0.72	4.3E-11	3.1E-03	0.71	0.17
223065_s_at	STARD3NL	-1.2	5.8E-06	3.7E-03	0.61	0.52
201206_s_at	RRBP1	-0.99	2.4E-05	0.02	0.65	0.67
204158_s_at	TCIRG1	-1.21	2.5E-04	0.02	0.73	1
217861_s_at	PREB	-0.71	1.7E-07	0.02	0.69	0.58
204683_at	ICAM2	0.81	5.0E-03	0.02	0.78	0.80
200644_at	MARCKSL1	-1.49	3.9E-03	0.04	0.66	1
202369_s_at	TRAM2	-0.55	3.7E-08	0.05	0.69	1

T cells

Probe id	Symbol	log2FC	Ref_p	Dif_p	R^2	CC
224252_s_at	FXYD5	-0.34	5.9E-03	0.02	0.17	0.17
215346_at	CD40	0.28	0.02	0.03	0.25	0.25
209496_at	RARRES2	0.34	0.07	0.04	0.25	0.25

B cells

Probe id	Symbol	log2FC	Ref_p	Dif_p	R^2	CC
222538_s_at	APPL1	-2.58	8.8E-05	4.E-06	0.61	0.89
202539_s_at	HMGCR	-0.89	1.2E-09	4.E-05	0.77	1
204552_at	INPP4A	0.86	0.01	3.E-03	0.65	1
200971_s_at	SERP1	0.74	4.0E-03	5.E-03	0.66	0.56
204678_s_at	KCNK1	-0.64	4.1E-07	5.E-03	0.61	0.88
204912_at	IL10RA	0.6	2.2E-04	6.E-03	0.67	1
210785_s_at	THEMIS2	0.85	0.02	6.E-03	0.68	1
212712_at	CAMSAP1	-0.89	9.33E-05	8.E-03	0.64	1
220306_at	FAM46C	0.83	0.02	0.01	0.70	1
212324_s_at	VPS13D	-1	0	0.02	0.65	0.89
200866_s_at	PSAP	0.71	0.01	0.02	0.64	0.28
1554252_a_at	CERS3	-1.65	0.02	0.02	0.73	1
206896_s_at	GNG7	0.78	0.04	0.04	0.62	1

Supplemental Figure 1. Flowchart of the PSEA computational steps