Single-cell RNA-seq uncovers cellular heterogeneity and provides a signature for paediatric sleep apnoea

Rene Cortese[1], Taylor S. Adams[2], Kylie H. Cataldo[1], Justin Hummel[3], Naftali Kaminski[2], Leila Kheirandish-Gozal[1] and David Gozal[1]

[1]Department of Child Health, Child Health Research Institute, School of Medicine, University of Missouri, Columbia, MO, USA. [2]Pulmonary, Critical Care and Sleep Medicine, Yale School of Medicine, New Haven, CT, USA. [3]Institute for Data Science and Informatics, University of Missouri, Columbia, MO, USA.

Corresponding author: David Gozal (gozald@health.missouri.edu)

Shareable abstract (@ERSpublications)
This study used scRNA-seq technology to investigate for the first time the fluctuation of cellular populations in PBMCs of OSA children and combine this information to generate a molecular biomarker signature for OSA diagnosis https://bit.ly/3ftGUTq

Cite this article as: Cortese R, Adams TS, Cataldo KH, *et al*. Single-cell RNA-seq uncovers cellular heterogeneity and provides a signature for paediatric sleep apnoea. *Eur Respir J* 2023; 61: 2201465 [DOI: 10.1183/13993003.01465-2022].

This article has an editorial commentary: https://doi.org/10.1183/13993003.02316-2022

Received: 25 July 2022
Accepted: 22 Sept 2022

Abstract

Background Obstructive sleep apnoea (OSA) is a highly prevalent disease and a major cause of systemic inflammation leading to neurocognitive, behavioural, metabolic and cardiovascular dysfunction in children and adults. However, the impact of OSA on the heterogeneity of circulating immune cells remains to be determined.

Methods We applied single-cell transcriptomics analysis (scRNA-seq) to identify OSA-induced changes in transcriptional landscape in peripheral blood mononuclear cell (PBMC) composition, which uncovered severity-dependent differences in several cell lineages. Furthermore, a machine-learning approach was used to combine scRNAs-seq cell-specific markers with those differentially expressed in OSA.

Results scRNA-seq demonstrated OSA-induced heterogeneity in cellular composition and enabled the identification of previously undescribed cell types in PBMCs. We identified a molecular signature consisting of 32 genes, which distinguished OSA patients from various controls with high precision (area under the curve 0.96) and accuracy (93% positive predictive value and 95% negative predictive value) in an independent PBMC bulk RNA expression dataset.

Conclusion OSA deregulates systemic immune function and displays a molecular signature that can be assessed in standard cellular RNA without the need for pre-analytical cell separation, thereby making the assay amenable to application in a molecular diagnostic setting.

Introduction

Obstructive sleep apnoea (OSA) is the most severe phenotype of sleep disordered breathing (SDB) and is highly prevalent across all age groups with an estimated prevalence of 2–10% in the general population [1]. Importantly, OSA affects 1.2–5.7% of children, with a peak prevalence occurring at 2–8 years of age [2]. Many of the clinical characteristics of paediatric OSA, and the determinants of its epidemiology, differ from those of adult OSA [3]. However, similarly to adults, OSA is a major cause of cardiovascular morbidity [4] and neurocognitive dysfunction [5] in children.

Polysomnography is the gold standard for the diagnosis of SDB in adults and children [6], although it represents a significant burden for the patient, insurers and the health system at large. There is a need for molecular diagnostic methods which can simplify OSA diagnosis, and provide an accurate, quick and affordable method, which is applicable in a clinical setting [7]. Several groups explored different biomarker types with uneven accuracy results. A systematic review on OSA biomarkers [8] summarised that so far only a few studies reported sensitivity and specificity values for the proposed biomarkers, which varied substantially from 43% to 100%, and inconsistencies in the biomarker performances arose from sample collection and processing protocols [9].

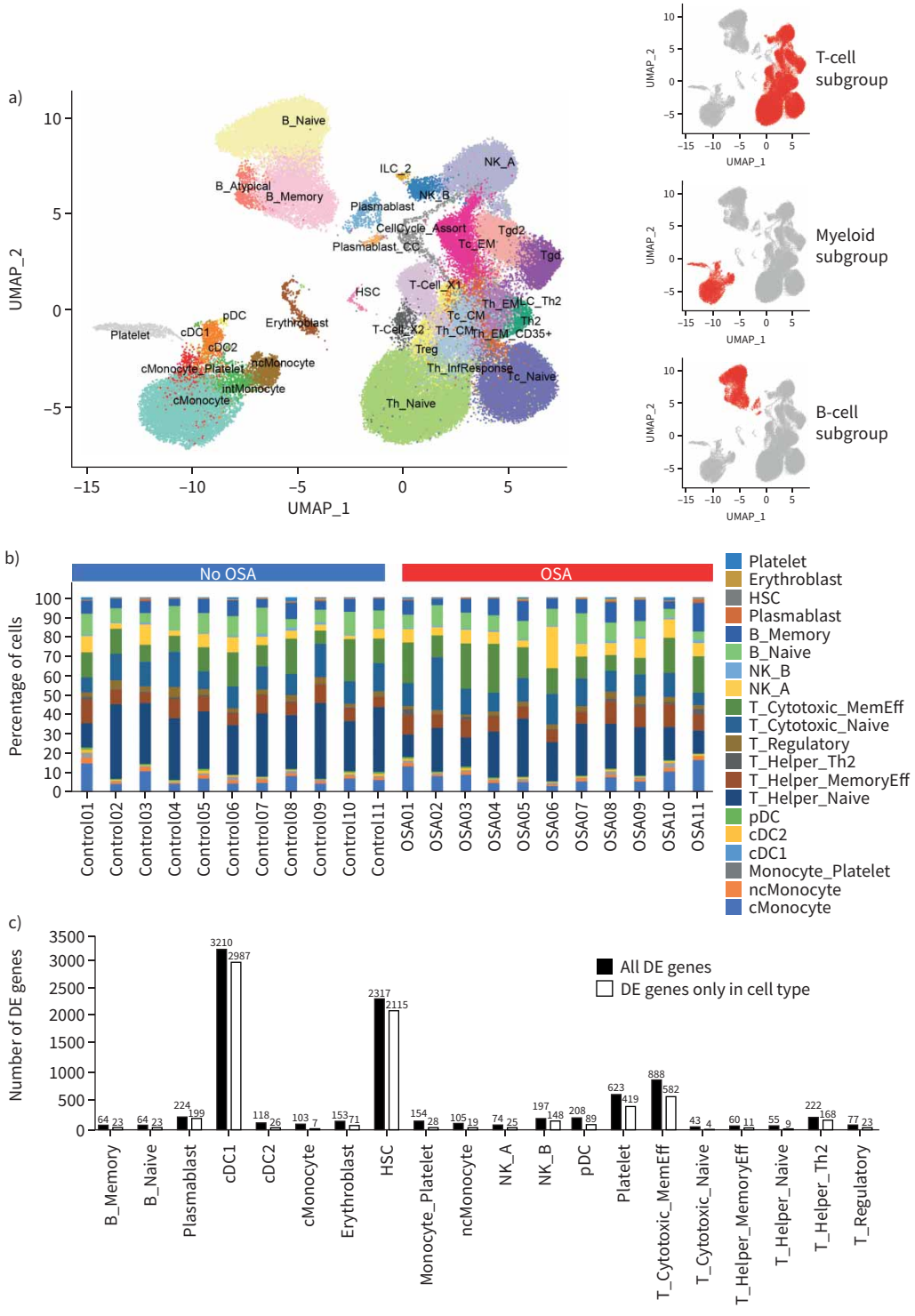

FIGURE 1 Identification of cell populations among the peripheral blood mononuclear cells (PBMCs) by single-cell RNA sequencing (scRNA-seq). a) Uniform manifold approximation and projection (UMAP) plotting enabled the identification of PBMC populations based solely on their scRNA-seq profiles. The three cellular lineages in PBMCs clustered separately: myeloid, T-cell and B-cell subgroups, respectively. b) The percentage of cells for each cell population varied in each subject in the obstructive sleep apnoea (OSA) and control groups. Colour-coded staked bars represent the percentage of cells for each cell population on the total number of cells in each subject. c) Differentially expressed (DE) genes in each PBMC population. Whereas some genes were differentially expressed in more than one cell type, some genes were unique for each population. Black bars indicate the number of genes that are differentially expressed in each PMBC population; white bars

indicate the number of genes that are differentially expressed only in each PBMC population. B_Memory: memory B-cells; B_Naive: naive B-cells; cDC: conventional dendritic cells; cMonocyte: classical monocytes; erythroblast: erythroblast cells; HSC: heamatopoietic stem cells; monocyte_platelet: monocyte expression features with platelets; ncMonocyte: nonclassical monocytes; NK: natural killer cells; pDC: plasmacytoid dendritic cells; T_Cytotoxic_MemEff: memory effector cytotoxic T-cells; T_Cytotoxic_Naive: naive cytotoxic T-cells; T_Helper_MemEff: memory effector helper T-cells; T_Helper_Naive: naive T-helper cells; T_Helper_Th2: T-helper type 2 cells; T_Regulatory: regulatory T-cells.

The impact of OSA on the transcriptional landscape of circulating immune cells in children and adults remains to be determined. Due to the low-invasive methods for sample collection and its extensive use in clinical practice, peripheral blood mononuclear cells (PBMCs) constitute a preferred source for biomarker discovery and development of diagnostic assays. A meta-analysis showed potential for OSA biomarkers in blood of paediatric patients, with insulin, 8-isoprostane and α-amylase among them [10]. PBMCs showed differential gene expression in healthy adult volunteers exposed to intermittent hypoxia [11] and in OSA patients compared to controls [12, 13]. A bulk RNA transcriptomic analysis in nonobese children with OSA previously identified inflammatory pathways as the driving network set of genes induced by the presence of the disease [14]. Furthermore, OSA-associated phenotypic features are associated with the altered expression of genes and proteins that distinguish cellular populations in PBMCs [15]. Tan *et al.* [16] demonstrated that paediatric OSA induces a shift to a pro-inflammatory state, which is associated with reduced T-regulatory cell lymphocytic population and altered T-helper (Th)1:Th2 balance toward Th1 predominance. Lastly, de la Peña *et al.* [17] showed that OSA patients have lower levels of endothelial progenitor cells (EPCs), and reduced expression of CD34 and vascular endothelial growth factor receptor 2, while we have shown that the level of recruitment of EPCs to the circulation is associated with vascular dysfunction in children with OSA [18].

In this study, we took advantage of the power of single-cell RNA sequencing (scRNA-seq) technology to investigate for the first time the fluctuation of cellular populations in PBMCs of OSA children and combine this information to generate a molecular biomarker signature for OSA diagnosis.

Methods

Patient population and sample preparation

The OSA group consisted of 11 patients who were polysomnographically diagnosed with OSA (apnoea–hypopnoea index (AHI) $\geqslant$5 events·h^{-1} of sleep). The control group consisted of 11 asymptomatic individuals without OSA (AHI <1 events·h^{-1}). OSA patients and controls were matched by age (mean±SD ages 6.64±1.27 years and 5.73±0.59 years for the OSA and control groups, respectively; p=0.528, t-test). Compared with controls, children in the OSA group had slightly higher body mass index (BMI) (mean±SD BMI 22.71±2.72 kg·m^{-2} and 17.92±1.95 kg·m^{-2} for the OSA and control groups, respectively; p=0.056, t-test) and significantly elevated AHI (mean±SD AHI 16.65±2.68 events·h^{-1} and 1.25±0.13 events·h^{-1} for the OSA and control groups, respectively; p=1.85×10^{-4}, t-test), in agreement with previously reported criteria [19]. All the participants provided written informed consent and the research protocol was approved by the research ethics board at the University of Missouri (protocol ID MU_2016683). PBMCs were isolated by gradient centrifugation using Ficoll-PAQUE (GE Healthcare Pharmacia, Chicago, IL, USA), cell viability was verified, and samples were cryopreserved in liquid nitrogen until use.

Single-cell transcriptomics analysis

Details on scRNA-seq methodologies and data analytics are provided as supplementary material. In brief, quality control of the isolated single cells was conducted thoroughly to monitor the suitability of each sample. scRNA-seq libraries were generated from isolated PBMCs using Chromium 3′ v3.1 (10X Genomics, Pleasanton, CA, USA). Libraries were sequenced using a NextSeq (Illumina, San Diego, CA, USA) instrument. The dataset is available at the National Center for Biotechnology Information Gene Expression Omnibus repository (accession number GSE214865). Data were processed using CellRanger (version 3.0.2). Raw reads were assessed by FastQC, followed by trimming of adapters using TrimGalore, and mapping using STAR. All data analysis was conducted in R (version 4.1.0). Cluster analysis was performed using the R package Seurat (version 4.0.5) [20]. Significant differences in cell type compositions were assessed by Wilcoxon rank-sum test. Correlations analyses were performed by Spearman correlation test. The percentage makeup was visualised *via* loess regression (95% CI). Principal component analysis and variable plots were conducted using the factoextra package version 1.0.7. Potential overlaps with biological processes and pathways were determined using Ingenuity Pathway Analysis software (Qiagen, Valencia, CA, USA).

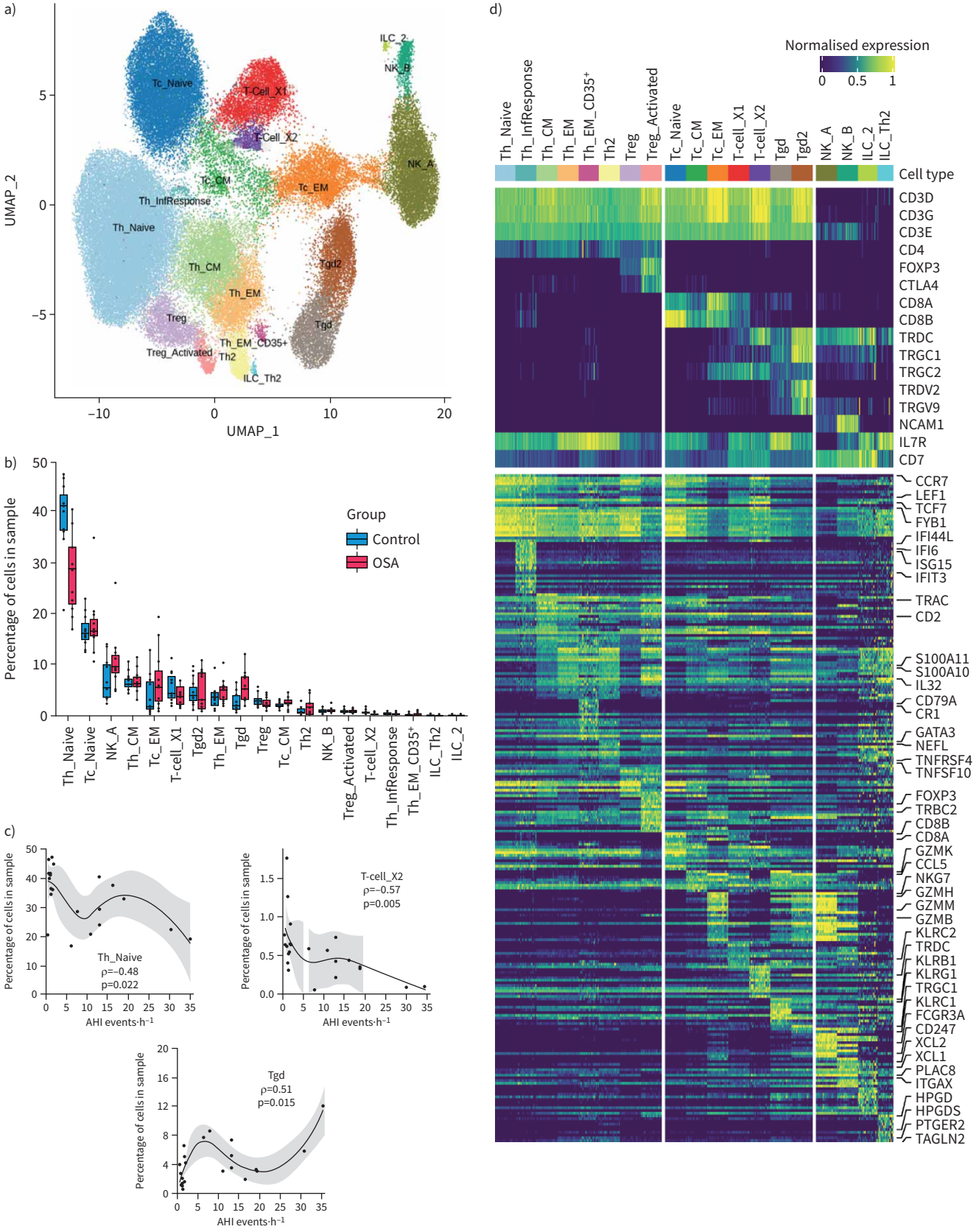

FIGURE 2 Cellular heterogeneity in the T-cell subgroup. a) The T-cell subgroup (n=76 646 cells) contained cell clusters corresponding to T-helper (Th), regulatory T (Treg), cytotoxic T (Tc), Tγδ (Tgd), natural killer (NK) and innate lymphoid (ILC) cells. Furthermore, we identified two cell clusters

whose transcriptional profiles did not correspond to any canonical T-cell population (*i.e.* T-Cell_X1 and T-Cell_X2). b) Obstructive sleep apnoea (OSA) patients showed a significant decrease in the number of naive T-helper (Th_Naive) cells compared with controls (averaged cell percentage 28.35±2.41% and 39.62±2.23% for the OSA and control groups, respectively; p=0.002, Wilcoxon test). c) Correlation between patient apnoea–hypopnoea index (AHI) and the percentage makeup of T-cells. OSA patients showed a significant decrease in the number of Th_Naive cells (ρ= −0.48, p=0.022) and the noncanonical X2 cell population (T-Cell_X2; ρ= −0.57, p=0.005), as well as an increase in Tgd cells (ρ=0.51, p=0.015). Percentage of the number of cells and AHI values are represented in the y- and x-axes, respectively. Shaded areas represent 95% confidence intervals. Correlations were considered significant when p<0.05 (Spearman correlation test). d) Heatmap of differentially expressed (DE) genes across cell clusters in the T-cell subgroup. T-cell populations and marker genes are accommodated in columns and rows, respectively. The top panel contains classical markers for T-cells; the lower panel contains the top markers (sorted by log 2 fold change) for each cell population. The average expression value in each cell population is colour-coded using a gradient from blue over green to yellow (average expression 0, 0.5 and 1, respectively). Tc_Naive: cytotoxic naive T-cells; Th_CM: central memory T-cells; Tc_EM: effector memory cytotoxic T-cells; T-Cell_X1: noncanonical T cells 1; Th_EM: effector memory helper T-cells; Tc_CM: central memory cytotoxic T-cells; Treg_Activated: activated regulatory T-cells; T-Cell_X2: noncanonical T-cells 2; Th_InfResponse: interferon-activated helper T-cells; Th_EM_CD35^{+}: effector memory CD35-positive helper T-cells.

Building of molecular signature

The molecular signature was extracted from each of the Seurat objects corresponding to specific cell type and clinical attributes, and combined, according to sample ID using the stringr (version 1.4.0) package. Empirical, binormal and nonparametric statistical methods, each using randomForest (version 4.7.1), were the parameters leveraged in the construction of three receiver operating characteristic (ROC) curves using the R package ROCit (version 2.1.1). Training and testing datasets were created using a 70:30 split to generate the observed ROC curves for each statistical method. An additional ROC-area under the curve (AUC) analysis was performed to validate the observed results using an independent microarray-based gene expression dataset [14]. Implementation of the ROC analysis was conducted using the same computational parameters outlined for building the signature. The independent dataset consisted of 1) OSA patients (n=28); 2) primary snoring individuals (children who snore, *i.e.* have symptoms of SDB, but their sleep studies are within normal limits) (n= 28); and 3) no OSA/no snoring controls (n=20). Implementation of the ROC analysis was conducted using the molecular signature with the same computational parameters, outlined for the following groups listed: 1) normal *versus* OSA; 2) normal *versus* primary snoring; and 3) primary snoring *versus* OSA.

Results

Single-cell transcriptional profiling discriminates cellular populations in PBMCs from OSA children and controls

Uniform manifold approximation and projection demonstrated that cells clustered according to their expression profiles (figure 1a). Cellular composition varied in OSA and control subjects (figure 1b and supplementary table S1), and when cells were stratified according to OSA status (supplementary figure S1a) and OSA severity (supplementary figure S1b). It is noteworthy that single-cell expression profiles were heterogeneous across individuals (supplementary figure S1c), highlighting the potential for personalisation of the marker panels. Figure 1c shows the number of differentially expressed genes in each cell type.

Cellular heterogeneity in the T-cells subgroup

The "T-cells subgroup" (n=76 646 cells) contained cell clusters corresponding to previously described T-cell populations (figure 2a). All identified subpopulations distinctively expressed known markers and were well represented across the samples. OSA patients exhibited a decline in the percentage of T-helper naive cells (Th_Naive; median±SD 28.35±2.41% and 39.62±2.23% for the OSA and control groups, respectively; p=0.002, Wilcoxon test) (figure 2b). Furthermore, we identified two cell clusters whose transcriptional profiles did not correspond to any canonical T-cell population (*i.e.* T-cell_X1 and T-cell_X2). AHI was significantly inversely correlated with the percentage of cells in the T-cell_X2 (ρ= −0.57, p=0.005; Spearman correlation test) and Th_Naive B (ρ= −0.48, p=0.022; Spearman correlation test) groups. Furthermore, AHI was significantly directly correlated with the percentage of cells in the γδ T-cell group (ρ=0.51, p=0.015; Spearman correlation test) and marginally significantly directly correlated with the percentage of cells in the NK_B-cell group (ρ=0.42, p=0.053; Spearman correlation test) (figure 2c).

Differentially expressed genes between the different clusters (figure 2d and supplementary table S2) included genes whose expression is associated with known T-cell populations (*e.g. CD3*, *CD4*, *FOXP3*, *CTL4*, *etc.*) as well as other genes which may define previously unknown cellular populations that are characteristic among OSA patients. We observed putative activation of molecular networks specific for the

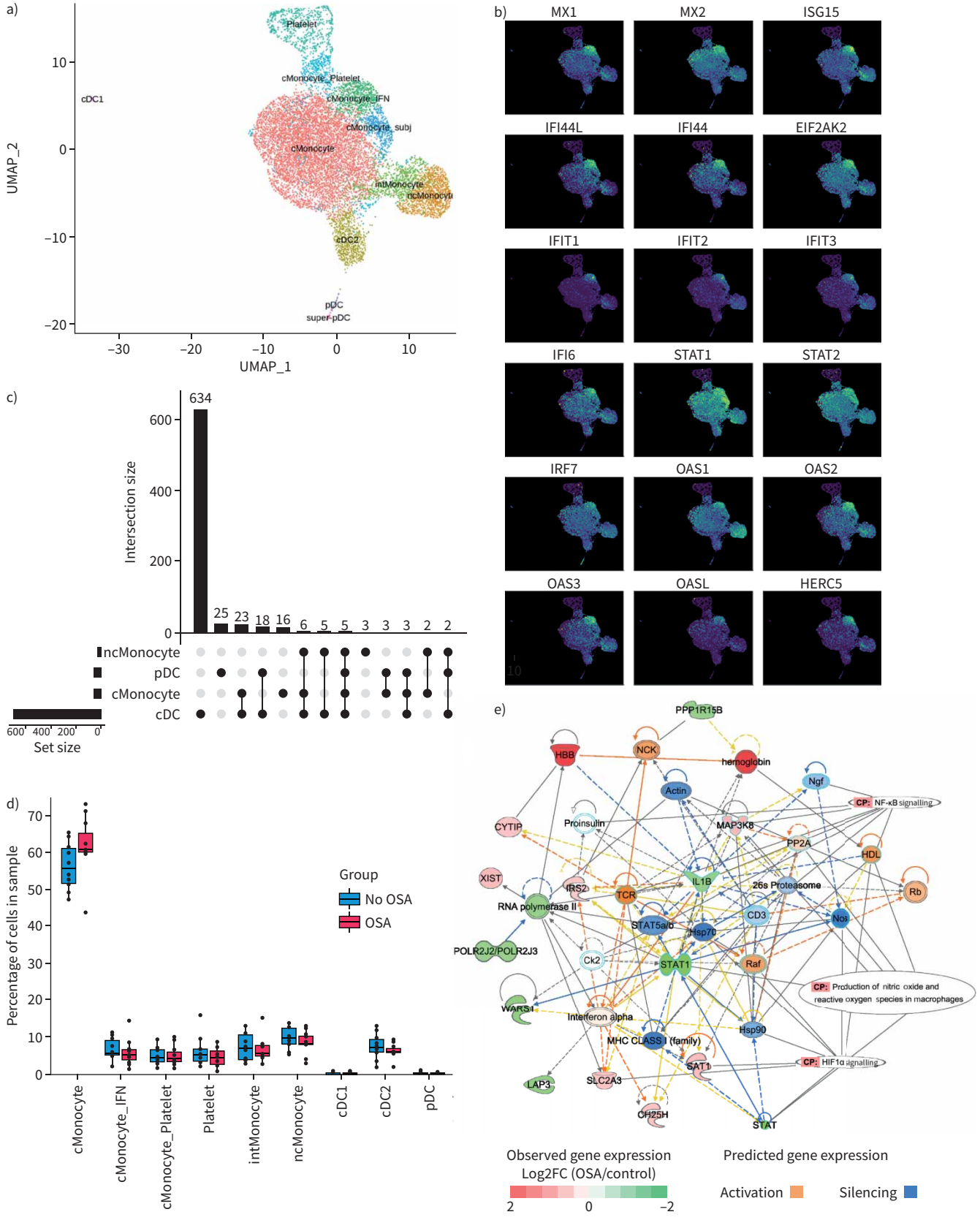

FIGURE 3 Cellular heterogeneity in the myeloid subgroup. a) The myeloid subgroup (n=11 708 cells) contained cell clusters corresponding to canonical myeloid cellular populations (*i.e.* classical monocytes (cMonocytes), non-classical monocytes (ncMonocytes), platelets, conventional

dendritic cells (cDCs) and plasmacytoid dendritic cells (pDCs)), as well as clusters sharing cMonocyte gene expression features with platelets (cMonocyte_Platelet) or ncMonocytes (classical monocytes expressing features with non-classical monocytes (intMonocytes)) and a cluster (Monocyte_IFN) whose gene expression profiles corresponded to an interferon (IFN)-primed phenotype. b) The Monocyte_IFN group expressed markers of an IFN-primed phenotype. Cells expressing each marker are coloured using a gradient based on the expression value from purple over green to yellow (average expression 0, 0.5 and 1, respectively). c) UpSet plots showing the number of genes differentially expressed in one or more myeloid cellular populations. Out of 895 differentially expressed (DE) genes, 674 were differentially expressed uniquely in one of the cell types: 16 out of 58 in cMonocytes, three out of 23 in ncMonocytes, 634 out of 741 in cDC and 25 out of 56 in pDC groups. d) Obstructive sleep apnoea (OSA) patients showed a significant increase in the percentage of cMonocytes compared with controls (average cell percentage±SD 63.72±1.62% and 56.49±2.02% for the OSA and control groups, respectively; p=0.035, Wilcoxon test). The OSA and control groups are represented as in figure 2. e) Analysis of molecular networks revealed cell-type specific regulation induced by OSA in cMonocytes. DE genes in the cMonocytes group were associated with "production of nitric-oxide and reactive oxygen species", as well as "hypoxia-inducible factor (HIF)1α signalling" and "NF-κB signalling". Reported and predicted interaction between the genes is represented by solid and dashed lines, respectively. The observed expression of each gene is colour-coded using a gradient from red to green (high expression in OSA and control groups, respectively). Genes with predicted activation or inhibition are coloured in orange and blue, respectively.

T-Cell_X1 and T-Cell_X2 populations that are associated with hypoxia-inducible factor signalling and oxidative stress, two main cellular processes in OSA pathophysiology (supplementary figure S2, supplementary table S3).

Cellular heterogeneity in the myeloid subgroup

Gene expression profiles in the "myeloid subgroup" (n=11 708 cells) enabled the identification of cellular clusters corresponding to known myeloid cell populations (figure 3a). Remarkably, we identified a cluster (Monocyte_IFN) whose gene expression profiles faithfully corresponded to an interferon (IFN)-primed phenotype (figure 3b). We identified 895 genes showing differential expression between the monocyte types, of which 674 were differentially expressed uniquely in one of the cell types (figure 3c, supplementary table S4).

The rate of cMonocyte cells was elevated in OSA patients compared with controls (average cell percentage±SD 63.72±1.62% and 56.49±2.02% for the OSA and control groups, respectively; p=0.035, Wilcoxon test), whereas cellular composition of the other clusters was equivalent (figure 3d). Analysis of molecular pathways and networks revealed cell-type-specific regulation induced by OSA in myeloid cells (figure 3e, supplementary table S5). Differentially expressed genes in the cMonocytes group were associated with "production of nitric-oxide and reactive oxygen species", "hypoxia-inducible factor (HIF) 1α signalling", and "NF-κB signalling".

Cellular heterogeneity in the B-cells subgroup

Single-cell expression profiles distinguished cluster of cells in the "B-cells subgroup" (n=17 010 cells) (figure 4a). The list of genes expressed in each B-cell cluster is provided in supplementary table S6. Subpopulations within the B_Memory cells were identified according to class-switch recombination (CSR) status [21] (supplementary figure S3). In addition, we identified two clusters of plasma B-cells (Plasmablast and Plasmablast_CC) which have undergone CSR (supplementary figure S4a) and express immunoglobulin at higher rates than other B-cell populations (supplementary figure S4b and c). One of these populations (Plasmablast_CC) expressed high levels of transcripts related to mitosis and DNA synthesis (supplementary figure S4) highlighting a high rate of cell proliferation. The B_Atypical population consisted of a mixed cell population of $CD18^{+}/CD20^{+}/CD21^{-}/CD27^{-}$ B-cells sharing a distinctive gene expression signature (supplementary figure S5a–c). Atypical B-cells could be also characterised according to their CSR status (supplementary figure S5d and e). Remarkably, the rate of cells belonging to B_Atypical subpopulation was significantly higher (coefficient of variation=82%) in OSA samples compared with the control group (average cell percentage±SD 4.15±1.07% and 2.28±0.68% for OSA and control groups, respectively; p=0.013, Wilcoxon test) (figure 4b).

Principal component analysis (PCA) including counts for the B-cell subgroups and phenotypic variables (*i.e.* age, BMI and AHI) showed that OSA patients clustered separately from controls (figure 4c). The AHI and subpopulations of mature B-cells (*i.e.* B_Memory, B_Atypical, Plasmablast and Plasmablast_CC) were the variables dragging the sample distribution towards the OSA samples. Conversely, B_Naive counts dragged sample distribution towards the position of control samples (figure 4d). In contrast, similar PCA analysis using counts from the T-cells and myeloid subgroups discriminated the OSA and control groups to a much lesser extent (supplementary figure S6a and b).

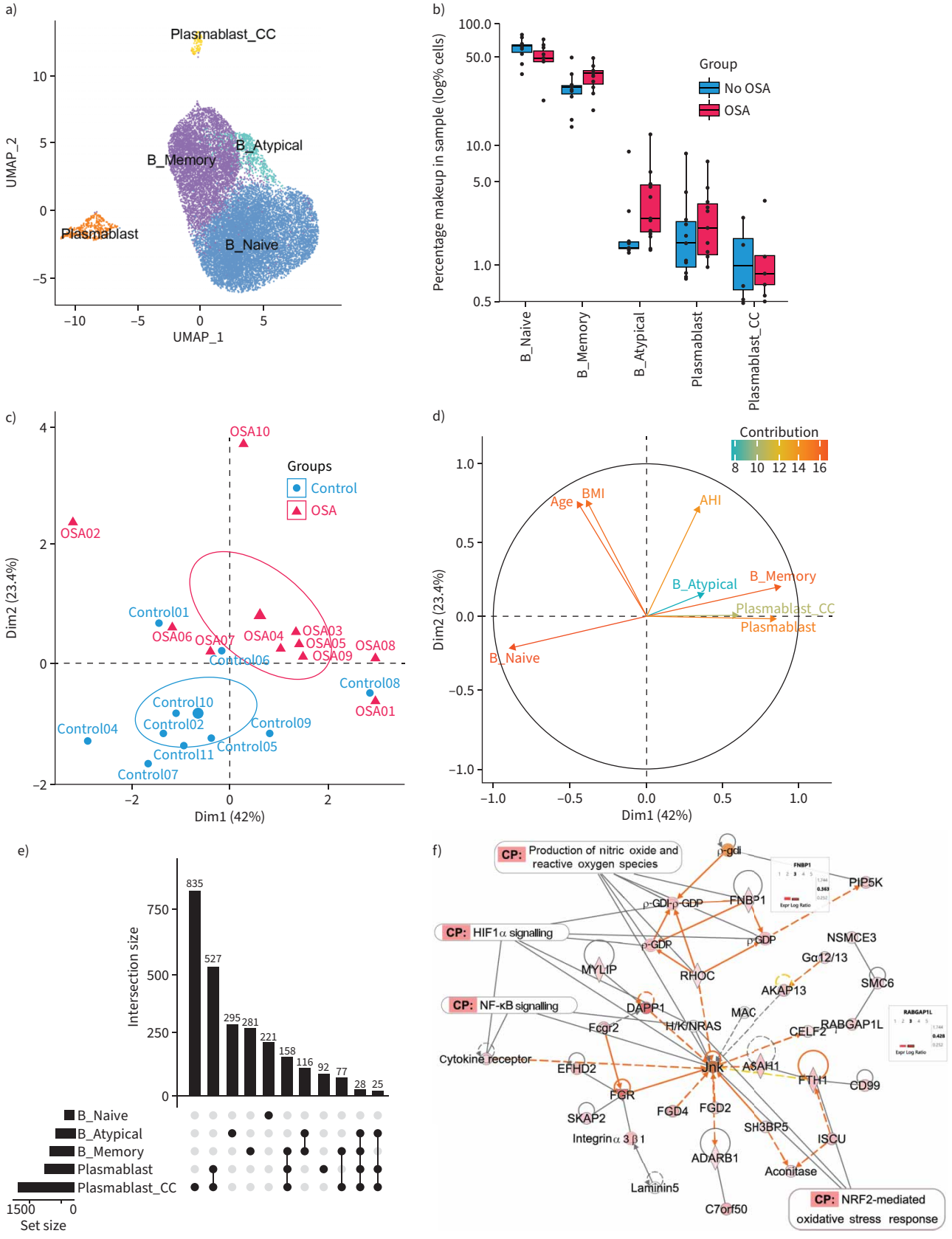

FIGURE 4 Cellular heterogeneity in the B-cell subgroup. a) The B-cell subgroup (n=17 010 cells) contained clusters corresponding to B_Naive, B_Memory, Plasmablast and B_Atypical populations. b) Obstructive sleep apnoea (OSA) patients exhibited a significant increase in the percentage

of B_Atypical cells compared with controls (average cell percentage 4.15±1.07% and 2.28±0.68% for the OSA and control groups, respectively; p=0.013, Wilcoxon test), as well as a marginally significant increase in B_Memory cells (average cell percentage: 39.34±3.18% and 31.39±3.50% for OSA and control groups, respectively; p=0.088, Wilcoxon test) and decrease in B_Naive cells (average cell percentage: 56.51±3.99% and 66.33±3.60% for OSA and control groups, respectively; p=0.076, Wilcoxon test) in OSA patients compared with controls. The percentage of cells in each cell population is presented on a log scale (y-axis) to highlight differences between populations with low percentage makeup. c) Principal component analysis (PCA) plot using the percentage of each B-cell population, and phenotypic variables (*i.e.* apnoea–hypopnoea index (AHI), age and body mass index (BMI)) distinguishes between OSA patients as well as controls (red and blue shapes, respectively). d) Graphic of variables indicating the direction and contribution for each variable in the PCA analysis from panel c). Positively and negatively correlated variables point to the same or opposite side of the plot, respectively. The contributions to the sample discrimination are colour-coded in a gradient from red (higher contribution) over yellow to light blue (lower contribution). e) UpSet plots showing the number of genes differentially expressed (DE) in one or more B-cell populations. Out of 4010 DE genes, 1726 were differentially expressed in only one cell type: 221 out of 256 in B_Naive, 281 out of 700 in B_Memory, 295 out of 520 in B_Atypical, 92 out of 858 in Plasmablasts and 835 out of 1676 in plasmablast cells expressing cell cycle features (Plasmablast_CC). f) Example of an OSA-activated molecular network in B-cells. Gene interactions and expression are represented as in figure 3. The expression values (exp log ratio) of two genes differentially expressed across B-cell populations are highlighted in the inserted bar plots. B-cell populations are encoded with numbers: 1=B_Naive, 2=B_Memory, 3=B_Atypical, 4=Plasmablast and 5=Plasmablast_CC.

We identified 4010 genes that were differentially expressed across the B-cell subgroups, with 1726 being differentially expressed in only one cell type (figure 4e). Differentially expressed genes in all B-cell types were related to canonical pathways associated with B-cell development, activation and immune-cell crosstalk, although differences were noted among the different B-cell subpopulations (supplementary table S7). Comparative analysis of molecular networks involved processes related to OSA pathophysiology also reflected distinctive transcriptional profiling for each B-cell subpopulation. Noteworthy were the identified activated networks in the two populations that were increased in OSA patients (*i.e.* B_Atypical and B_Memory groups; figure 4g and f, respectively), in which expressed genes were associated with biological processes and mechanisms activated in OSA pathophysiology, such as “oxidative stress”, “HIF1α” and “NF-κB signalling”. These results are in line with those in the T-cell and myeloid subgroups and highlight the changes in overall cellular heterogeneity induced by paediatric OSA leading to cell-specific functional pathway dysregulation in PBMCs.

Building and validation of a molecular signature for diagnosis of paediatric sleep apnoea

We applied a machine-learning approach to build a molecular signature that discriminates between OSA and control subjects by combining markers of PBMC cellular composition with those associated with the occurrence of OSA. A molecular signature consisting of 32 genes (figure 5a) was extracted from each of the Seurat objects. A list of the genes in the signature with its corresponding annotation is provided in supplementary table S8. The performance of the signature was evaluated using three statistical methods for distinguishing between OSA and control subjects in the same sample set and a cross-validation strategy. Training and testing datasets were created to generate the observed ROC curves for each statistical method (figure 5b). Results for the empirical (AUC 0.96), binormal (AUC 0.96) and nonparametric ROC (AUC 0.95) demonstrated that the molecular signature has a very high power to discriminate between patients and controls, and therefore warranted further validation studies.

Next, we performed additional ROC-AUC analyses to assess the performance of the molecular signature in a bulk RNA expression microarray-based independent PBMC sample set [14]. In addition, the performance of the signature was further evaluated according to positive predictive values (PPV) and negative predictive values (NPV). The molecular signature showed a high performance for discriminating OSA patients from no OSA/no snoring controls, with the ROC-AUC analysis resulting in AUCs of 0.93, 0.96 and 0.92 for empirical, binormal and nonparametric ROCs, respectively (figure 5c), with 93% and 95% PPV and NPV, respectively. Likewise, the molecular signature had a high performance to distinguish between OSA patients from primary snoring individuals (figure 5d) (AUCs 0.97, 0.99 and 0.96 for empirical, binormal and nonparametric ROCs, respectively; 96% PPV and 95% NPV) and primary snoring individuals from no OSA/no snoring controls (AUC 0.95, 0.98 and 0.95 for empirical, binormal and nonparametric ROCs, respectively; 96% PPV and 95% NPV).

Discussion

Single-cell transcriptomic profiling is gaining momentum and has been extensively used to study cellular heterogeneity in several diseases and tissue types [22]. We and others have studied transcriptomic profiles in PBMC from OSA patients using “bulk” RNA approaches, such as gene expression microarrays [13, 14, 23] or RNA-seq [24]. Notwithstanding, the current findings illustrate the heretofore unknown OSA-induced cellular mosaicism in peripheral blood cells and their associated transcriptional landscape, providing very

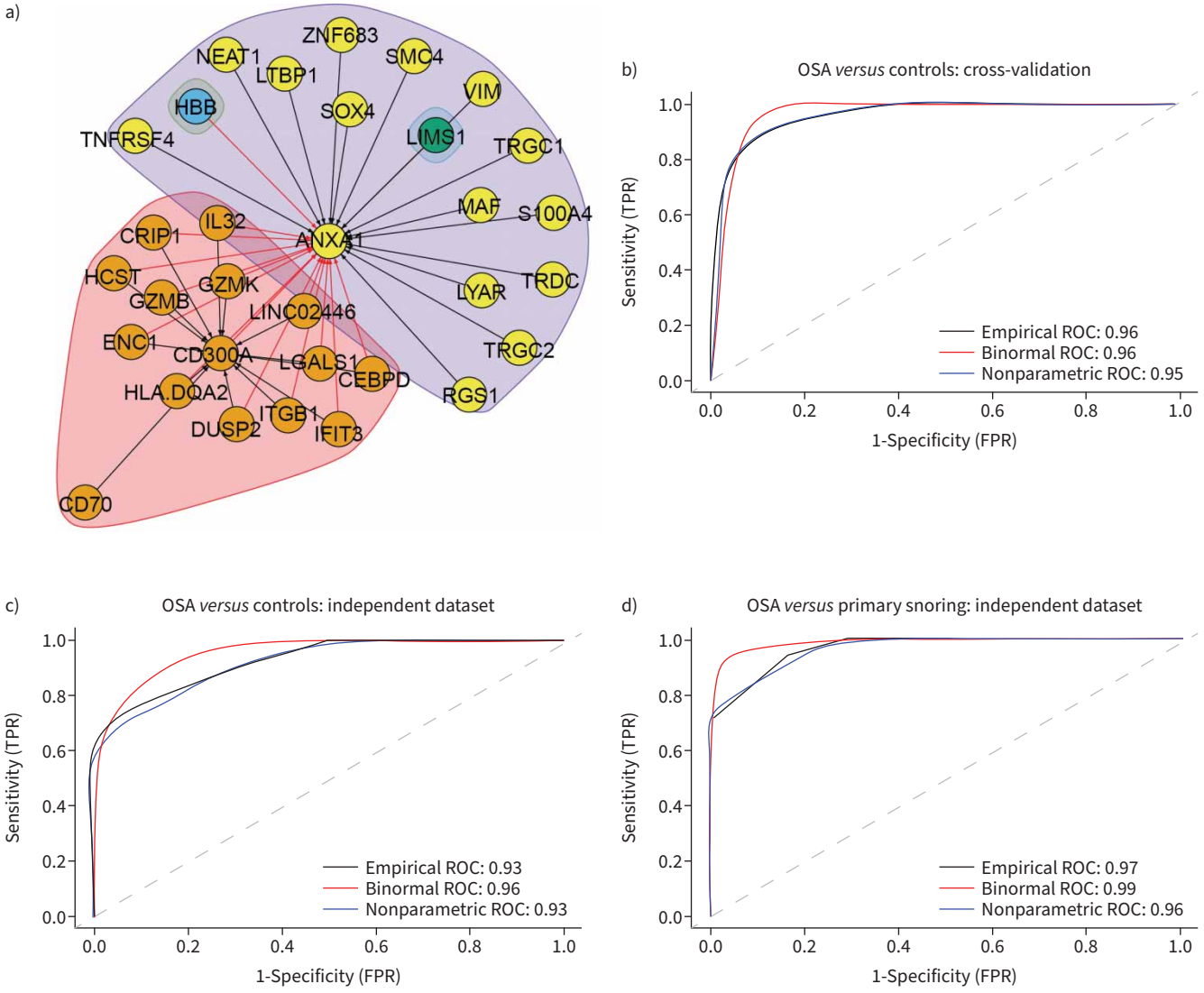

FIGURE 5 Molecular signature for the diagnosis of paediatric sleep apnoea. a) Molecular signature consisting of 32 markers. Connecting lines depict interactions between markers. Shaded areas represent markers showing higher interaction in the molecular network. The colour of each node represents the weight of the node in the molecular network. b) Receiver operating curve (ROC) analysis assessing the performance of the molecular signature for the discrimination between obstructive sleep apnoea (OSA) and control subjects during the cross-validation strategy. The performance of the signature was assessed by calculating the area under the curve (AUC) when applying three different models: empirical (AUC 0.96), binormal (AUC 0.96) and nonparametric (AUC 0.95), represented as black, red and blue lines, respectively. c) ROC analysis assessing the performance of the molecular signature for the discrimination between OSA and control subjects using an independent dataset [14]. AUC results for the empirical (AUC 0.93), binormal (AUC 0.96) and nonparametric (AUC 0.93) modelling are presented as in panel b). d) ROC analysis assessing the performance of the molecular signature for the discrimination between OSA and primary snoring subjects using an independent dataset [14]. AUC results for the empirical (AUC 0.97), binormal (AUC 0.99) and nonparametric (AUC 0.96) modelling are presented as in panel b). TPR: true positive rate; FPR: false positive rate.

novel insights for understanding the impact of paediatric OSA on immune cell populations. Importantly, the study of single-cell transcriptomic profiles enabled the identification of previously undescribed PBMC cell types and the building of a molecular signature for high-precision molecular diagnosis of OSA.

OSA-induced changes in cellular heterogeneity occurs in the three cell lineages in PBMCs (*i.e.* T-cells, myeloid cells and B-cells). We identified two previously undescribed T-cell subpopulations (*i.e.* T-Cell_X1 and T-Cell_X2), whose expression profiles were mainly related to T-cell activation, inflammation and intracellular signalling, as well as two major pathways in OSA: HIF1α signalling and oxidative stress [25]. Specific cellular signalling pathways are activated in the T-Cell_X2 population (*i.e.* phospholipase-C,

hepatocyte growth factor and WNT/β-catenin signalling pathways) compared with other T-cell populations. These pathways were previously reported as triggered by intermittent hypoxia, a hallmark of OSA [26]. Intermittent hypoxia leads to increased oxidative stress and activation of pro-inflammatory pathways, induces an elevation of the sympathetic tone and causes alterations of cardiovascular and metabolic parameters and pro-oxidative cellular phenotypes [27], as well as increase in oxidative stress and inflammation [28, 29]. In addition, we identified a monocyte subpopulation whose expression profile corresponded to an IFN-primed phenotype (*i.e.* monocyte_IFN). This is a discrete variety of classical monocytes with a stereotypic transcriptional phenotype, whose relative abundance has been linked to inflammation and promoting T-cell response in tumours [30]. Furthermore, the percentage of classical monocytes was significantly increased in OSA subjects compared to controls. Lastly, we detected an atypical B memory cell subpopulation (*i.e.* B_Atypical) which was significantly increased in OSA samples compared with the control group. Remarkably, molecular networks associated with oxidative stress and HIF and NF-κB signalling were activated in these B-cell types, as observed in the T-cell and myeloid populations. Taken together, these findings show the role of molecular mechanisms related to inflammation and oxidative stress, which ultimately will define unique cellular phenotypes in PBMCs as a response to OSA and its physiological components. Although more research is needed to fully identify the dynamics of the cellular populations, our findings support the hypothesis of OSA eliciting changes in cellular heterogeneity in PBMC, defining new cellular populations and holding the potential of markers detected by scRNA-seq for defining patient endotypes [6–9].

Large-scale profiles have been applied successfully as sources for candidate biomarkers, but their application as molecular diagnostics is hampered by costs and other intrinsic limitations (*e.g.* complex pre-analytical procedures, generation of information that is beyond the clinical application, *etc.*). Here, we applied a novel machine-learning approach that combines the power of scRNA-seq (*i.e.* cell-specific markers) with traditional gene expression analysis (*i.e.* differentially expressed genes associated with OSA patients). With this approach, we identified a molecular signature consisting of 32 genes, which can distinguish OSA patients from controls with high precision and high PPV and NPV in an independent microarray-based "bulk" RNA expression dataset. Note that ROCs showed equivalent performances in both populations (*e.g.* AUCs for empirical ROCs were 0.96 and 0.93 for the cross-validation and independent dataset validation, respectively). Very importantly, datasets were very different: 1) samples were isolated from different subjects; 2) RNA expression was assessed by two different methods (scRNA-seq in cross-validation (figure 5b) and microarray-based in independent validation (figure 5c); and 3) gene expression was quantified using single-cell (cross-validation; figure 5b) or "bulk" (independent dataset validation; figure 5c) RNA approaches. Despite all these differences, the molecular signature was able to robustly identify OSA subjects from controls with high precision and high PPV and NPV, demonstrating the reproducibility of the biomarker panel as well as the applicability of such panel in standard RNA samples in clinical settings. Future studies will formally assess the performance of this signature in the context of cell-specificity, to characterise the involvement of relevant morbidity-related pathophysiological mechanisms and to confirm the clinical utility of the molecular signature in larger and more heterogeneous sample sets.

Acknowledgements: The authors would like to thank Wes Warren and Ashley Meyer (University of Missouri, Columbia, MO, USA) for technical support during the scRNA-seq sample processing.

Data availability: De-identified data for this study, including scRNA-seq profiles, will be available immediately following publication with no end date to anyone who wishes to access the data. The dataset is available at the NCBI Gene Expression Omnibus (GEO) repository (accession number GSE214865).

Author contributions: R. Cortese participated in the conceptual framework of the project, performed experiments, analysed data, and drafted components of the manuscript. K.H. Cataldo performed experiments. T.S. Adams and J. Hummel conducted data analysis. N. Kaminski and L. Kheirandish-Gozal provided expert guidance, participated in data analysis and interpretation of results, and served as blinded observers. D. Gozal conceptualised the project, provided critical input in all phases of the experiments, analysed data, drafted the ulterior versions of the manuscript, and is responsible for the financial support of the project and the manuscript content. All authors have reviewed and approved the final version of the manuscript.

Conflict of interests: All authors have nothing to disclose.

Support statement: This work was partially supported by a Leda J. Sears Charitable Trust Grant and a BioNexus KC 21-1 Patton Trust Research Grant to R. Cortese. D. Gozal is supported by NIH grant AG061824 and by Tier 2 and TRIUMPH grants from the University of Missouri. Funding information for this article has been deposited with the Crossref Funder Registry.

References

1 Benjafield AV, Ayas NT, Eastwood PR, *et al.* Estimation of the global prevalence and burden of obstructive sleep apnoea: a literature-based analysis. *Lancet Respir Med* 2019; 7: 687–698.
2 DelRosso LM. Epidemiology and diagnosis of pediatric obstructive sleep apnea. *Curr Probl Pediatr Adolesc Health Care* 2016; 46: 2–6.
3 Gozal D. Sleep, sleep disorders and inflammation in children. *Sleep Med* 2009; 10: Suppl. 1, S12–S16.
4 Vrints H, Shivalkar B, Hilde H, *et al.* Cardiovascular mechanisms and consequences of obstructive sleep apnoea. *Acta Clin Belg* 2013; 68: 169–178.
5 Kheirandish-Gozal L, Yoder K, Kulkarni R, *et al.* Preliminary functional MRI neural correlates of executive functioning and empathy in children with obstructive sleep apnea. *Sleep* 2014; 37: 587–592.
6 Khurana S, Soda N, Shiddiky MJA, *et al.* Current and future strategies for diagnostic and management of obstructive sleep apnea. *Expert Rev Mol Diagn* 2021; 21: 1287–1301.
7 Byrnes SA, Weigl BH. Selecting analytical biomarkers for diagnostic applications: a first principles approach. *Expert Rev Mol Diagn* 2018; 18: 19–26.
8 de Luca Canto G, Pachêco-Pereira C, Aydinoz S, *et al.* Biomarkers associated with obstructive sleep apnea: a scoping review. *Sleep Med Rev* 2015; 23: 28–45.
9 Bencharit S, Redenz RG, Brody ER, *et al.* Salivary biomarkers associated with obstructive sleep apnea: a systematic review. *Expert Rev Mol Diagn* 2021; 21: 223–233.
10 Kheirandish-Gozal L, Gozal D. Pediatric OSA syndrome morbidity biomarkers: the hunt is finally on! *Chest* 2017; 151: 500–506.
11 Polotsky VY, Bevans-Fonti S, Grigoryev DN, *et al.* Intermittent hypoxia alters gene expression in peripheral blood mononuclear cells of healthy volunteers. *PLoS One* 2015; 10: e0144725.
12 Chen YC, Chen KD, Su MC, *et al.* Genome-wide gene expression array identifies novel genes related to disease severity and excessive daytime sleepiness in patients with obstructive sleep apnea. *PLoS One* 2017; 12: e0176575.
13 Lin SW, Tsai CN, Lee YS, *et al.* Gene expression profiles in peripheral blood mononuclear cells of Asian obstructive sleep apnea patients. *Biomed J* 2014; 37: 60–70.
14 Khalyfa A, Capdevila OS, Buazza MO, *et al.* Genome-wide gene expression profiling in children with non-obese obstructive sleep apnea. *Sleep Med* 2009; 10: 75–86.
15 Kheirandish-Gozal L, Wang Y, Duggan RC, *et al.* Nitric oxide production by monocytes in children with OSA and endothelial dysfunction. *Clin Sci* 2014; 127: 323–330.
16 Tan H-L, Gozal D, Wang Y, *et al.* Alterations in circulating T-cell lymphocyte populations in children with obstructive sleep apnea. *Sleep* 2013; 36: 913–922.
17 de la Peña M, Barceló A, Barbe F, *et al.* Endothelial function and circulating endothelial progenitor cells in patients with sleep apnea syndrome. *Respiration* 2008; 76: 28–32.
18 Kheirandish-Gozal L, Bhattacharjee R, Kim J, *et al.* Endothelial progenitor cells and vascular dysfunction in children with obstructive sleep apnea. *Am J Respir Crit Care Med* 2010; 182: 92–97.
19 Kang K-T, Lee P-L, Weng W-C, *et al.* Body weight status and obstructive sleep apnea in children. *Int J Obes* 2012; 36: 920–924.
20 Butler A, Hoffman P, Smibert P, *et al.* Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* 2018; 36: 411–420.
21 Stavnezer J, Guikema JEJ, Schrader CE. Mechanism and regulation of class switch recombination. *Annu Rev Immunol* 2008; 26: 261–292.
22 Aldridge S, Teichmann SA. Single cell transcriptomics comes of age. *Nat Commun* 2020; 11: 4307.
23 Fan C, Huang S, Xiang C, *et al.* Identification of key genes and immune infiltration modulated by CPAP in obstructive sleep apnea by integrated bioinformatics analysis. *PLoS One* 2021; 16: e0255708.
24 Turnbull CD, Lee LYW, Starkey T, *et al.* Transcriptomics identify a unique intermittent hypoxia-mediated profile in obstructive sleep apnea. *Am J Respir Crit Care Med* 2020; 201: 247–250.
25 Cofta S, Winiarska HM, Płóciniczak A, *et al.* Oxidative stress markers and severity of obstructive sleep apnea. *Adv Exp Med Biol* 2019; 1222: 27–35.
26 Yuan G, Nanduri J, Khan S, *et al.* Induction of HIF-1α expression by intermittent hypoxia: involvement of NADPH oxidase, Ca^{2+} signaling, prolyl hydroxylases, and mTOR. *J Cell Physiol* 2008; 217: 674–685.
27 Gileles-Hillel A, Almendros I, Khalyfa A, *et al.* Early intermittent hypoxia induces proatherogenic changes in aortic wall macrophages in a murine model of obstructive sleep apnea. *Am J Respir Crit Care Med* 2014; 190: 958–961.
28 Jun J, Savransky V, Nanayakkara A, *et al.* Intermittent hypoxia has organ-specific effects on oxidative stress. *Am J Physiol Regul Integr Comp Physiol* 2008; 295: R1274–R1281.
29 Gozal D, Kheirandish-Gozal L. Cardiovascular morbidity in obstructive sleep apnea: oxidative stress, inflammation, and much more. *Am J Respir Crit Care Med* 2008; 177: 369–375.
30 Mulder K, Patel AA, Kong WT, *et al.* Cross-tissue single-cell landscape of human monocytes and macrophages in health and disease. *Immunity* 2021; 54: 1883–1900.