The global transcriptomic signature in sinonasal tissues reveals roles for tissue type and chronic rhinosinusitis disease phenotype*

Ahmed Bassiouni[1], Judy Ou[1], Andreas Schreiber[2,3,4], Joel Geoghegan[2], Anna Tsykin[2], Alkis J. Psaltis[1], Peter-John Wormald[1], Sarah Vreugde[1]

[1] Department of Otolaryngology, University of Adelaide, Adelaide, Australia

[2] Australian Cancer Research Foundation (ACRF) Genomics Facility, Centre for Cancer Biology, SA Pathology, Adelaide, Australia

[3] School of Biological Sciences, University of Adelaide, Adelaide, Australia

[4] School of Pharmacy and Medical Science, University of South Australia, Adelaide, Australia

Rhinology 58: 3, 273 - 283, 2020

https://doi.org/10.4193/Rhin19.403

***Received for publication:**

November 3, 2019

Accepted: December 30, 2019

Abstract

Background: RNA sequencing (RNA-Seq) allows the characterization of a global transcriptomic signature in a least-biased fashion, but few studies have applied this method to investigate the pathophysiology of CRS.

Methods: We collected mucosal tissue samples from 6 CRS without nasal polyps (CRSsNP), 6 CRS with nasal polyps (CRSwNP), and 6 control patients. Additional matched polyp samples were collected from the 6 CRSwNP patients. RNA was extracted and sequenced on the Illumina HiSeq-2500. Differential gene expression and pathway analyses were performed.

Results: CRSsNP showed evidence of upregulated interferon-mediated immunity, MHC-class-I mediated antigen presentation, CXCR3 binding, neutrophil chemotaxis and degranulation, and potential downregulation of genes related to cilia movement and production. CRSwNP polyp tissue showed upregulation of B-cell mediated immune responses, but reduced expression of genes related to epithelial morphogenesis and haemostasis. Polyps also showed a generalized reduction of positive gene regulation. The sinonasal transcriptomic signature was largely determined by tissue type (polyp versus mucosa) and disease phenotype, with minimal signal originating from the individual patient.

Conclusion: RNA-Seq is a useful tool to explore the complex pathophysiology of CRS. Our findings stress the importance of tissue selection in molecular research utilizing sinonasal tissue, and demonstrate the limitation of the sNP/wNP paradigm (and the importance of endotyping). On the other hand, classical CRSsNP/wNP disease phenotypes played some role in determining the global transcriptomic signature, and should not be hastily discarded. The value of RNA-Seq-described transcriptomic signatures in exploring endotypes is yet to be explored in future studies.

Key words: chronic rhinosinusitis, chronic rhinosinusitis with nasal polyps, CRSwNP, nasal polyposis, nasal polyps, RNA-Seq, mRNA, gene expression profiling, transcriptome, transcriptomic signature

Introduction

The constant and rapid advances in genomic science and in high-throughput sequencing technologies offer a unique opportunity to gain insights into the mechanisms of complex disease. Chronic Rhinosinusitis (CRS) is a complex upper airway inflammatory condition of unclear etiology[(1)] and pathology akin to asthma[(2)]. CRS is known to include more than one phenotype (CRS sine nasal polyps "CRSsNP", and with nasal polyps "CRSwNP")[(1)] and multiple endotypes [(3)]; a fact that complicates research into and reasoning about its pathophysiology. Consequently, despite years of CRS research, the etio-pathogenesis of CRS remains unknown, and steroids and nasal rinsing remain the cornerstone of medical treatment. Although new biologic treatments are being developed and offer great promise [(4)], no major breakthrough has been observed to date that would change current standard treatment protocols. As such many

patients remain with limited conservative options, and surgical treatment becomes necessary in these patients who fail medical treatment. Some patients even remain refractory to multiple surgical procedures, and these patients exhibit poor quality of life (QoL) and put a significant burden on the healthcare system [5].

These facts call for exploiting next-generation techniques that have the ability to produce "big data", which include genomics (whole genome sequencing "WGS"), proteomics (mass spectrometry), or transcriptomics (RNA sequencing "RNA-Seq"). Before the advent of these methods, most CRS research projects that utilized tissue collected from patients have only advanced at a pace of a few genes (and/or proteins) investigated per study. Next-generation techniques complement those traditional mechanistic studies, and allow the investigation of dysfunction at scale, directly in diseased tissues collected from patients. Microarrays have already been utilized by several groups and offered an exciting avenue to investigate the gene expression profiles in CRS at high-throughput scales [6–13]. RNA-Seq has now supplanted microarrays in characterizing the global transcriptomic picture in a least-biased fashion [14,15].

We hypothesized that the global transcriptomic signatures may be sufficiently different (between different disease phenotypes and between patients), in such a way that the global transcriptomic signature could be used in the future to explore CRS endotypes and explore disease progression through comparisons of tissue in different stages of disease progression isolated from the same patients. Here, we utilize RNA-Seq to investigate global transcribed gene (mRNA) expression profiles from CRSsNP, CRSwNP and non-CRS controls. We study global gene expression patterns, perform differential gene expression (DGE), and use the data to perform clustering. We report that the type of tissue sampled and disease phenotype play a significant role in determining the global transcriptomic signature.

Material and methods

Patient inclusion criteria and sample collection

The Queen Elizabeth Hospital Human Research Ethics Committee approved the study (approval HREC/15/TQEH/12). All patients included in the study had approved storage of their tissue for the purpose of research and had signed written informed consents.

Inclusion criteria were patients aged 18 years or older, requiring endoscopic sinus surgery either for CRS, or for skullbase tumours (control patients) who had had tissue collected and stored in the Department of Otolaryngology's research tissue bank. Exclusion criteria included CRS with comorbid asthma or aspirin-exacerbated respiratory disease (AERD), in order to reduce the number of confounding factors. Patients indicated for pre-operative systemic (oral) steroids were excluded, but patients were included even if they continued to use their topical nasal steroid spray pre-operatively. All samples were collected from participating patients intra-operatively during endoscopic sinus surgery under endoscopic visualization. Mucosal samples were collected from the anterior ethmoidal region from the CRSsNP and control patients. For the CRSwNP patients, two types of tissue were obtained: from the anterior ethmoidal/middle meatal region where polyp tissue was collected (termed henceforth "polyp"), as well as from the adjacent area of the middle tubinate where mucosa/polypoidal mucosa was collected (termed henceforth "mucosa"). CRSwNP patients were only included if they had both polyp and mucosal tissue types available.

All samples were placed in Eppendorf tubes containing RNALater® (Ambion) and transported on ice to the laboratory where they were stored at -80°C for further processing. The four different groups are therefore (Control_mucosa, CRSsNP_mucosa, CRSwNP_mucosa, CRSwNP_polyp); and are termed as such and as the "main study groups" throughout the manuscript. All samples from all study groups were collected, handled, transported and stored in the tissue bank in a similar manner.

RNA extraction

Samples were then pulled out of -80°C for RNA extraction. Samples were thawed and using a sterile scalpel were cut into 80-100mg pieces. Samples were then homogenized in QIAZol® using stainless-steel bead beating in the Qiagen TissueLyser machine (2 minutes, at 20-30Hz). RNA was then extracted from the homogenates using the RNEasy Mini Kits, following the manufacturer's protocol, which includes a DNAase step. RNA concentration and quality were checked on the Experion RNA StdSens system (Bio-Rad Laboratories, Hercules, CA, USA). RNA was stored at -80°C until sequencing.

RNA sequencing

Messenger RNA library preparation and sequencing was done at the Australian Cancer Research Foundation (ACRF) Cancer Genomics Facility. Samples were transported to the facility on dry ice. RNA Quality control was performed on the Agilent Bioanalyzer 2100 (Agilent). Sample quality was assured by only including samples with a Bioanalyzer RIN score greater than 7. One microgram of total RNA was enriched for mRNA using the NEBNext Poly(A) mRNA Magnetic Isolation Module. The NEBNext Ultra RNA Library Prep Kit for Illumina was then used to prepare non-directional RNA libraries. Libraries were sequenced on the Illumina HiSeq 2500 (1x50bp), utilizing four lanes on the instrument. Libraries of samples from all four study groups were barcoded and distributed equally across the four lanes to avoid lane-specific bias effect. The study was planned such that the number of patients and samples included yielded an approximate depth of 30 million reads/sample at the end of sequencing. This depth of sequencing was chosen according to the "Standards, Guidelines and Best Practices for RNA-Seq V1.0" (The

ENCODE Consortium, June 2011).

Bioinformatics pipeline and statistical analysis

Transcript-level quantifications were produced from the raw fastq files using Salmon [16]. At first, Salmon (version 0.12.0) was used to construct a transcriptome search index from the GENCODE [17,18] release 91 human transcriptome annotation (file: "gencode.v29.transcripts.fa.gz"), using a k (k-mer length) parameter of 23. This k-mer length was selected as it showed the best mapping rates of reads to the transcriptome after investigating k-mer lengths of 19, 21, 23, as well as the default 31. Since features could also include pseudogenes in the transcriptome annotations, the usage of the term gene in this manuscript thereafter will be used as a connotation to an annotated feature (i.e. gene or pseudogene) on the GENCODE transcriptome. The term transcript will be used to refer to the different isoforms that map to one gene/feature.

Salmon was then used for fast, bias-aware RNA transcript quantifications against the built index using the default quasi-mapping approach and default parameters. Salmon quantification files were imported and feature/gene-level count estimates were inferred from transcript-level counts using the R package tximport [19]. Counts normalization and differential gene expression analysis was done using DESeq2 version 1.18.1 [20]. Genes with very low counts (≤ 10 reads across all samples) were excluded from the dataset before DESeq2 analysis. DESeq2's Likelihood Ratio Test (LRT) was used to test differentially expressed genes in any of the main four study groups simultaneously against a null model. Differentially expressed genes were selected with a significance cutoff level of 0.05 for optimizing the False Discovery Rate (FDR) independent filtering. Independent hypothesis weighting (IHW)(21) was used to increase power for multiple comparisons. DESeq2's Wald tests was used for pairwise detection of differentially expressed genes between each of the study groups and the resulting log2-fold-changes were shrunk using DESEq2's 'lfcShrink' function with appropriate contrast as argument. The R package StageR was used to correct for the p-values generated from these pairwise Wald tests through its two-step screening-confirmation methodology [22]. Further subgroup analysis in CRSwNP samples was performed using DESeq2's Wald tests, but controlling for the patient as a fixed variable in the model specification. The regularized log normalization (the rlog function in DESeq2) was applied to the gene counts and the rlog-transformed counts were used for downstream analyses e.g. distance matrix generation and ordination. Significant differentially-expressed genes with absolute fold-change (FC) ≥ 2 were chosen as a subset for input to most of these downstream analyses.

Over-representation analyses for Gene Ontology (GO) [23,24] and Reactome pathway database [25] were done using the goseq R package [26], with separate analyses [27] for upregulated and downregulated genes. P values from goseq were corrected using the Benjamini-Hochberg FDR correction [28], with significant over-representation cutoff at adjusted p cutoff of 0.05. Supervised classification and feature selection was done using the sparse Partial Lease Squares Discriminant Analysis (sPLS-DA) algorithm [29] implemented in the R package mixOmics [30]. The algorithm's performance was tuned on the rlog-transformed counts of the whole dataset according to the authors' sPLS-DA tutorial.

All frontend analysis work was done through the Jupyter notebook frontend [31] and utilizing the assistance of packages from R (R Foundation for Statistical Computing, Vienna, Austria) and the Scientific Python stack [32]. Euclidean distance matrices between the gene level expression profiles of the samples were done after applying the regularized log normalization (the rlog function in DESeq2). The implementation of Principal Components analysis (PCA) in the python package scikit-learn version 0.20.0 [33]. PCA was performed on the subset of genes that were significantly differentiated (adjusted p-value cutoff of 0.05) and having a FC ≥ 2. Multivariate analysis using Permutational Multivariate Analysis of Variance (PERMANOVA)[34] was done through the implementation in the R package 'vegan' [35]. PERMANOVA was performed using a model formula "distance matrix ~ tissue type + disease + patient_id" to investigate the effect of these three covariates on the distance matrix. The Mantel test was used to test for the correlation between the CRSwNP polyp and mucosa distance matrices. A Procrustes analysis was then performed to rotate these matrices to maximum similarity and a Procrustes test (PROTEST) was done to test whether deviations between the transformed matrices are significantly less than that expected by chance [36,37].

Table 1. Basic characteristics of the 18 patients included in the study.

Clinical variable	CRSsNP	CRSwNP	Control
Mean age (Range)	47.0 (18 - 73)	54.0 (21 - 76)	57.7 (34 - 83)
Sex "Male"	6	6	4
Mean number of previous surgeries (Range)	0.8 (0 - 3)	1.0 (0 - 3)	0.0 (0 - 0)
Asthma	0	0	0
Aspirin-exacerbated Respiratory disease (AERD)	0	0	0
Specific-IgE confirmed allergic status	0	1 patient tested "moderate for grass and pollen"	0

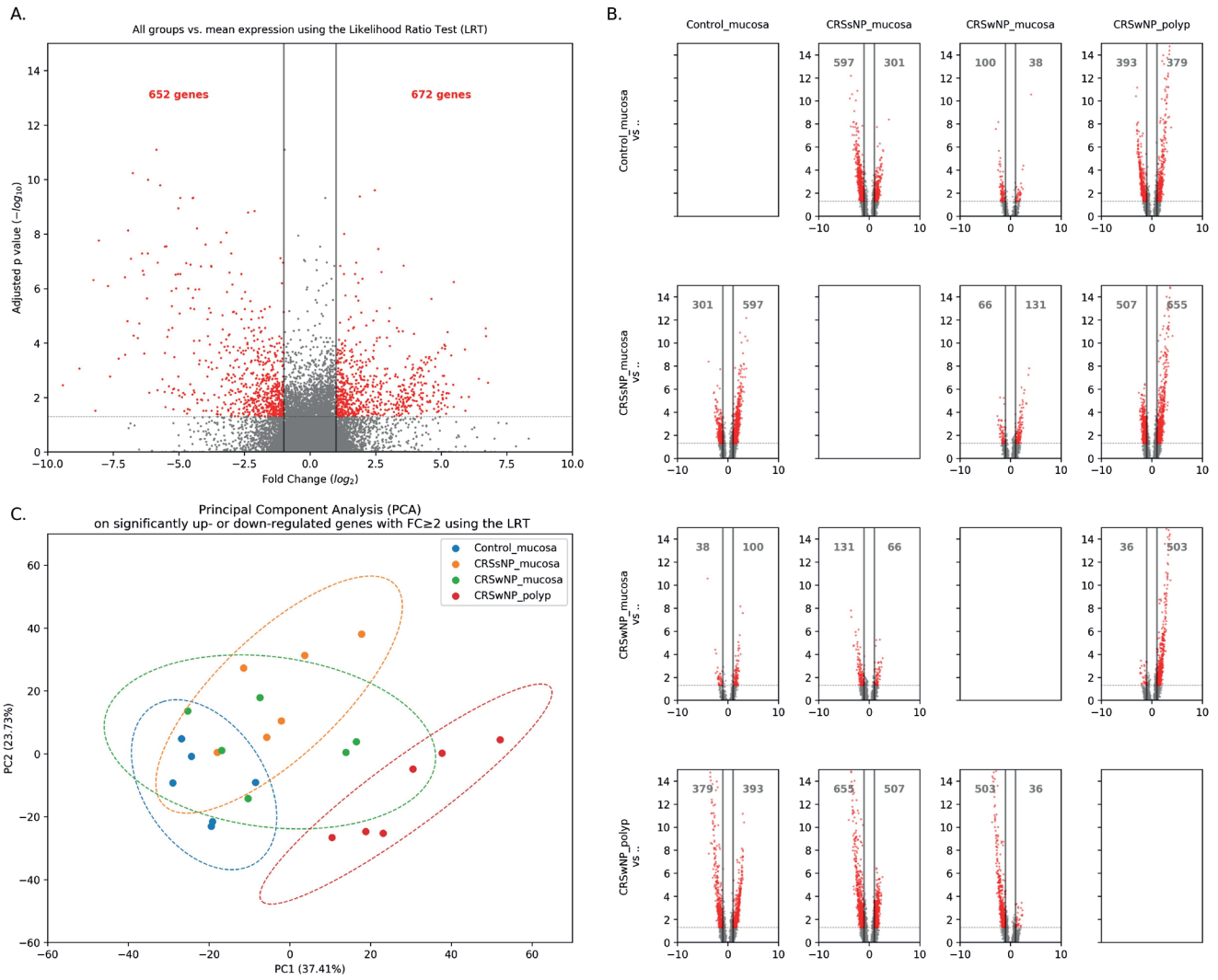

Figure 1. Differential gene expression results from DESeq2 analysis (A) Volcano plot highlighting the significantly up- and down-regulated genes (adjusted p < 0.05) detected using the Likelihood Ratio test (LRT) with absolute FC≥2. The volcano plot is a scatter plot where each gene is represented by a dot. The y-axis is -log10(p value) i.e. the higher the gene on the y-axis means the lower the p value. The x-axis is log2(FC) i.e. the left side of a volcano plot contains the down-regulated genes while the right side of the plot contains the up-regulated genes. (B) Matrix of volcano plots similar to (A) but demonstrating the pairwise differential expression analyses between each study group using Wald tests (with direction of comparison: groups labeled vertically on y-axis to left of figure, relative to groups labeled on x-axis horizontally on top of figure). (C) Principal Component Analysis (PCA) on the significantly expressed genes detected using LRT with absolute FC≥2. Each scatter point plotted represent one sample. Ellipses represent 95% confidence intervals of the co-ordinates of the samples in each group. FC=Fold change; PCA= Principal Component Analysis; PC1 = Principal Component 1; PC2 = Principal Component 2.

Results

Patients and samples

Twenty four samples were included, representing 18 patients and the four study main study groups, with six samples in each of the four study groups (Control_mucosa, CRSsNP_mucosa, CRSwNP_mucosa, and CRSwNP_polyp). The basic characteristics of the 18 patients included in the study is shown in Table 1.

Principal Component Analysis and PERMANOVA

We explored clustering of samples on a 2-dimensional plot using Principal Component Analysis (PCA) for dimensionality reduction. The regularized-log counts of the significantly differentiated genes on DESeq2's LRT with absolute fold-change ≥2 was used as the dataset fed to the PCA algorithm. The first two principal components (PCs) explained 37.41% and 23.73% of the variance in this dataset, respectively. Upon visual inspection of the plot (Figure 1), there was satisfactory separation (although not 100% discrete clustering) of the study groups, with the "polyp" tissue type samples from CRSwNP patients clustering particularly away from other samples. One sample from the CRS-

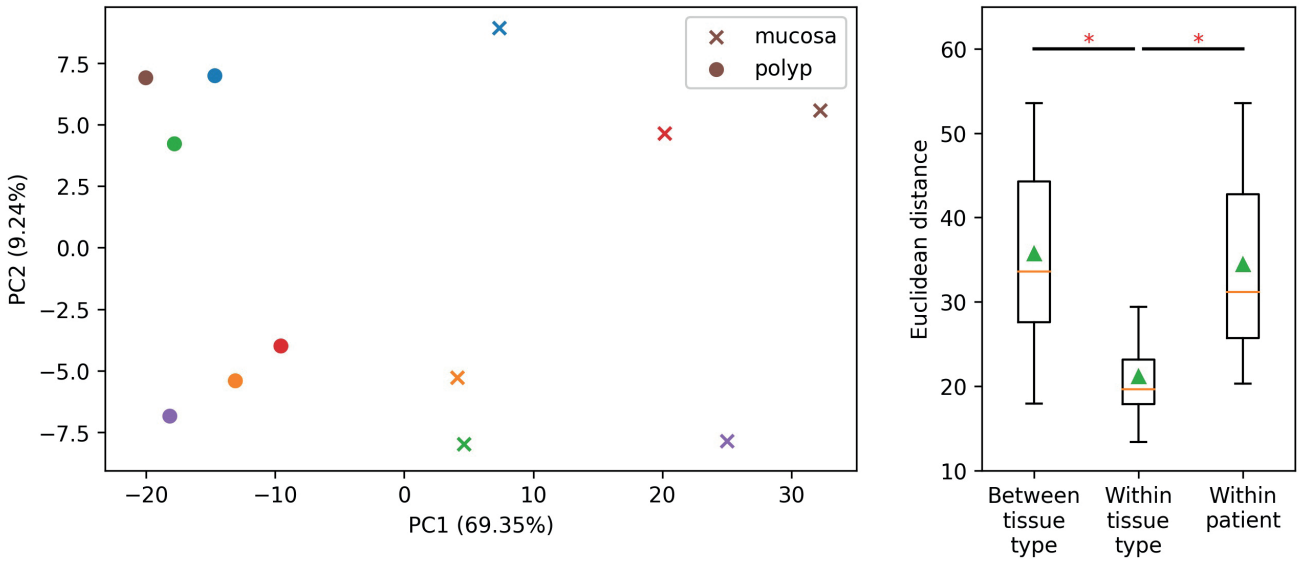

Figure 2. (A) Principal Component Analysis (PCA) of CRSwNP samples (=12) showing clustering by the type of tissue, not by patient. PCA was performed on the subset of significantly expressed genes with absolute FC≥2. (B) Boxplot illustrating the significant difference between the within-type versus the between-type and within-patient Euclidean distances ($p < 0.001$ and $p = 0.003$ respectively; Mann-Whitney test, with p values adjusted with the Benjamini-Hochberg FDR corrections). The boxes extend from the lower to upper quartile values and the whiskers show the range. The orange lines and green triangles represent the medians and means of the distances, respectively.

sNP group fell within the 95% confidence interval circle of the Control group. Mucosal tissue samples from CRSwNP patients appeared to cluster in-between the groups, showing a more variable placement between Control, CRSsNP and polyp tissue. To further investigate these observations, PERMANOVA[34] was performed to examine the effect of three covariates (tissue type, disease group, and patient) on the Euclidean distance matrix. PERMANOVA showed a significant effect of tissue type (pseudo-$F = 2.98$; $p = 0.001$) and a significant but lesser effect of disease group (pseudo-$F = 1.73$; $p = 0.005$). This indicates that the tissue type "polyp versus mucosa" had a more discernible global mRNA gene expression signature, when compared against the other three study groups. On the other hand, no significant effect was found for the patient variable (pseudo-$F = 1.09$; $p = 0.29$). This means that for these six CRSwNP patients that had two samples collected, the fact that the (two) samples were collected from the same patient did not appear to affect sample clustering. This finding highlights the crucial role of tissue sampling when investigating gene expression profiles in CRS studies.

Subgroup Analysis of CRSwNP patients: mucosa versus polyp comparison

We then performed a subgroup analysis on the matched mucosa and polyp samples collected from CRSwNP patients. To control for sample pairing (i.e. the effect of the patient variable), we added patient as a fixed effect in DESeq2's model specification and then we performed a DESeq2 Wald test DGE analysis to compare polyp tissue to mucosal tissue. This resulted in 122 upregulated genes and 511 downregulated genes (adjusted p value < 0.05); i.e. 80.73% of the differentially-expressed genes in polyp tissue were downregulated (rather than upregulated), when compared to mucosa tissue after control for sample pairing by patient. A repeat PCA for this dataset showed significant separation of the two tissue types on the first PC, which explained 69.35% of the variance of the statistically-significant differentially expressed genes with FC ≥ 2 (Figure 2A). Within-type distances were significantly less than within-patient and between-type distances (FDR-adjusted p values 0.003 and < 0.001 respectively; Mann-Whitney) (Figure 2B).

Multivariate analysis on the distance matrix using PERMANOVA showed no significant impact of the "patient" covariate (pseudo-$F = 1.13$; $p = 0.262$), in contrast to the tissue type covariate (pseudo-$F = 2.39$; $p = 0.004$).

The CRSwNP_mucosa and CRSwNP_polyp distance matrices were compared using the Mantel test, which tests the correlation between two matrices. This showed no significant correlation (Mantel $r = 0.498$; $p = 0.075$). In addition, Procrustes analysis and test (PROTEST) were done and showed that the sum of residual deviations between the matrices after Procrustes transformation was not significantly less than that expected by chance (Procrustes Sum of Squares = 0.076; $p = 0.09$).

Gene Ontology (GO) and Pathway analysis

To obtain a functional interpretation of the differentially expressed genes in our study groups, we performed a Gene Ontology (GO)[23] analysis and a pathway analysis (using the Re-

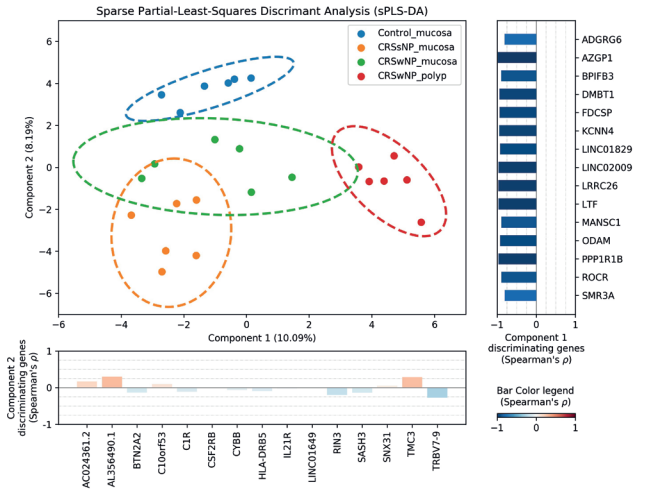

Figure 3. Sparse Partial Lease Squares Discriminant Analysis (sPLS-DA) was performed on the regularized-log "rlog" transformed counts of the whole dataset. Ellipses represent 95% confidence intervals of the co-ordinates of the samples in each group. Discriminant gene-sets are plotted along the axes with respective correlation coefficients to their respective Component.

actome pathways database [25]). The full results of these analyses can be found in the Supplementary Files.

To summarize these results:

- CRSsNP shows evidence of upregulated interferon-mediated immunity, including Type 1 interferon and interferon-gamma upregulation. (GO terms: GO:0060333, GO:0071346, GO:0060337, GO:0032729, GO:1902715, GO:0032609, GO:0034341, GO:0032727, GO:0045078 ; Reactome terms: R-HSA-1015702, R-HSA-1031716) This could be a marker of a Th1 response, or an antiviral response (GO terms: GO:0051607, GO:0009615, GO:0001618, GO:0046718, GO:0050690, GO:0019079; Reactome terms: R-HSA-1169410).
- CRSsNP shows over-representation of genes related to neutrophil chemotaxis (GO terms: GO:0030593, GO:0090023) and degranulation (GO:0043312).
- CRSsNP shows increased CXCR3 binding. (GO terms: GO:0048248; Reactome terms: R-HSA-374248).
- CRSsNP shows increased antigen presentation (GO GO:0003823, GO:0042605, GO:0015433; Reactome: R-HSA-2213244), specifically MHC class-I mediated with TCR complex (GO terms: GO:0042590, GO:0002476, GO:0002291, GO:0002480, GO:0002474), and Leukocyte immunoglobulin-like receptors (LILRs) interacting with MHC class-I molecules (Reactome term R-HSA-199043).
- CRSsNP shows activation of the complement cascade. This appears through the classical pathway, rather than the alternative pathway, and with particular participation of C4. (GO terms: GO:0006956, GO:0006958, GO:0030449; Reactome terms: R-HSA-166658, R-HSA-166753).
- CRSsNP shows evidence of a heightened innate immune reponse (GO terms: GO:0045087; Reactome terms: R-HSA-168249), which includes Natural Killer (NK) cell mediated cytotoxicity. (GO terms: GO:0030101, GO:0045954).
- CRSsNP shows a potential downregulation of ciliary movement and cilia component (e.g. axonemes) production pathways, especially when compared to CRSwNP. (GO terms: GO:0005930, GO:0005929, GO:0003341, GO:0045503, GO:0036159, GO:0031514, GO:0007018, GO:0060271, GO:0005874, GO:0036064, GO:0051959, GO:0060294, GO:0035082).
- CRSwNP polyp tissue shows reduced expression of genes related to epithelial morphogenesis. (GO terms: GO:0002009, GO:0061138).
- CRSwNP polyp tissue shows reduced degradation of extracellular matrix (GO terms: GO:0062023, GO:0090131, GO:0006508) and disturbed cell volume homeostasis (GO terms: GO:0006884).
- CRSwNP polyp tissue compared to control shows up-regulation of B-cell mediated immune responses e.g. positive regulation of B cell activation (GO:0050871) and B cell receptor signaling pathway (GO:0050853) with an increased antibacterial humoral response with increased production/activity of various immunoglobulins (GO terms: GO:0042571, GO:0034987, GO:0071748, GO:0071751, GO:0071752, GO:0071756) and increased immunoglobulin (Fc) receptor activity (GO:0034987, GO:0038095, GO:0038096).
- CRSwNP polyp tissue shows reduced expression of genes related to haemostasis (Reactome terms: R-HSA-109582) e.g. vasoconstriction (GO:0042310, GO:0014826), coagulation (GO:0007596) and platelet degranulation (GO:0002576).
- CRSwNP mucosa tissue shows upregulated C-C chemokine pathways (GO terms: GO:0070098, GO:0004950, GO:0016493, GO:0019957, GO:0019956). This included CCR1, CCR3 (which bind to eotaxins and RANTES - thereby promoting eosinophilic inflammation).
- CRSwNP polyp tissue shows generalized reduction of positive gene regulation (GO term: GO:0010628).

Discriminating gene-set selection using Discriminant Analysis

To identify a subset of genes that achieve maximum discrimination between the four study groups, we performed a type of Discriminant Analysis called sparse Partial Lease Squares Discriminant Analysis (sPLS-DA) [29]. sPLS-DA is a multivariate exploratory approach that is an extension of partial least squares and is capable of feature selection in multi-class classification problems [29]. The tuning of the sPLS-DA algorithm on our dataset recommended two components for classification/feature selection, with 15 discriminant genes on Component 1 and 15

Table 2. Statistics of gene up- and down-regulation in the four study groups. The direction of up- and down-regulation is as Group 1 compared to Group 2.

Group 1	Group 2	Up-regulated	Down-regulated	Upreg +FC2	Downreg +FC2
Control_mucosa	CRSsNP_mucosa	359	666	301	597
Control_mucosa	CRSwNP_mucosa	44	104	38	100
Control_mucosa	CRSwNP_polyp	441	447	379	393
CRSsNP_mucosa	CRSwNP_mucosa	164	81	131	66
CRSsNP_mucosa	CRSwNP_polyp	779	589	655	507
CRSwNP_mucosa	CRSwNP_polyp	566	65	503	36

Columns labeled "+FC2" include only genes up- or down-regulated with an absolute absolute fold change ≥2. FC = Fold Change.

discriminant genes on Component 2. Graphical results of the sPLS-DA are shown in Figure 3.

We note that CRSwNP_polyp samples separate from the other groups along Component 1 with significant downregulation of the selected Component 1 genes in the CRSwNP_polyp group. The Control_mucosa and CRSsNP_mucosa groups separate from each other along Component 2. The discriminant gene-sets selected by the algorithm are noted in the figure, with the correlation coefficients (Spearman's ρ) to their respective Component. Mucosal tissue from CRSwNP patients "CRSwNP_mucosa" showed a less distinctive pattern of gene expression among the selected gene-set, with samples from this group occupying an intermediate position between CRSsNP and CRSwNP polyp samples along the first and second components. (Figure 3) In general, we note that the graphical results from this semi-supervised approach are similar to the results of the unsupervised PCA ordination results (Figure 1C). The gene names with the mean counts for each gene in each study group is found in Supplementary Table S1.

Discussion

We explored global gene expression signatures in control and diseased sinonasal tissues using RNA-Seq. Our findings highlight a crucial role for the type of tissue, and a lesser role (albeit significant) for disease status on the transcriptomic signature. Moreover there appears to be, at least in our cohort, minimal signal originating from the individual patient, as indicated by absence of pairing of mucosa and polyp samples collected from the same patient.

Our study is not the first study to apply RNA sequencing (RNA-Seq) to sinonasal tissues isolated from human subjects [38–40]. However our study, to our knowledge, is the largest to date (n=24) and is the first to include paired mucosa and polyp samples collected from the same patients to explore insights into disease progression. Prior to the advance of RNA-Seq, its predecessor the microarray has been applied in numerous studies to study CRS [6–10,12,13].

CRSsNP shows evidence of upregulated interferon-mediated immunity, including Type 1 interferon and interferon-gamma upregulation. Interferon-gamma has been shown to be upregulated in a sinonasal mucosal explant model in response to Staphylococcus aureus biofilms [41]. CRSsNP showed evidence of neutrophil chemotaxis and degranulation and increased CXCR3 binding. CXCR3 tends to be expressed mainly on Th1 cells. CRSsNP also showed increased MHC class-I-biased antigen presentation. CRSsNP tissues showed a downregulation of genes involved in ciliary mechanisms, especially when compared to CRSwNP. This could be a contributor to the stagnation of mucus or "mucociliary disruption" in these patients and the need to re-establish muco-ciliary clearance with surgery, as explained in the famous works by MesserKlinger and Stammberger [42,43]. However, whether this downregulation in cilia-related genes reflects a consequence of the inflammation and the pathogenetic process in this disease, or whether it participates in disease initiation remains unknown and merits further investigation.

CRSwNP polyp tissue showed a severe depletion of lactotransferrin (alias lactoferrin; LTF). LTF is a component of innate immunity and has important antimicrobial and antibiofilm activity [44,45]. A reduced lactoferrin expression and activity has been reported in CRS tissue samples compared to controls and in sputum samples from cystic fibrosis patients [45–47]. Lactoferrin was also found to inhibit the proliferation of nasal polyp fibroblasts in vitro and thus reduced LTF could be a marker, or enhancer, of polyp growth and recurrence [48]. CRSwNP polyp tissue showed a severe depletion in DMBT1 gene expression (Deleted In Malignant Brain Tumors 1) [49]. DMBT1 (alias Salivary Agglutinin, Glycoprotein-340) is a protein-coding gene that has been reported to play various roles in the immune defense at mucosal surfaces and is considered a candidate tumor suppressor gene. It is thought to bind bacteria and bacterial products in a calcium-dependent way playing a role in bacterial defense via

NOD2 and TLR4 [50]. It is also thought to regulate the sensation of taste by binding proteins in saliva. Interestingly, taste receptors have recently been implicated in the innate sensing of bacterial products and immune defense against pathogens with polymorphisms associated with susceptibility to gram-negative upper respiratory infection [51]. Hence the role of the DMBT1 in the binding of bacterial products and how this may relate to the regulation of innate immune sensing and the potential role of taste receptors in this process remains an avenue of future investigation. A significant downregulation both LTF and DMBT1 in nasal polyps was reported in the microarray study by Rostkowska-Nadolska et al. [11]. On the other hand, these two genes were reported as significantly-overexpressed in polyps by Liu et al. [6], in a direct contradiction to our findings [6]. The explanation for this discrepancy is unknown but strengthens the argument for carefully selecting tissues collected for studying CRS pathogenesis in humans. The severe downregulation of Bpifb3 (BPI fold-containing family B member 3) in the CRSwNP_polyp group could be a contributor to the anosmia experienced in CRSwNP patients. This protein has been recently documented in the transcriptome and proteome of the main and acccessory olfactory epithelia in mice and thus could be an important gene in mammalian olfaction [52]. CRSwNP polyp tissue also showed reduced expression of genes related to haemostasis. This is an interesting finding and is in agreement with previous clinical literature. Nasal polyps were found in some studies to be a risk factor for a higher incidence/amount of intra-operative/peri-operative bleeding when compared to CRSsNP [53–56], with evidence that a higher polyp load (polyp score or Lund-Mackay score) was associated with a higher amount of blood loss [57,58]. GO analysis showed that CRSwNP polyp tissues, compared to controls, demonstrated an upregulation of B-cell mediated immune responses. This is in line with previous research indicating increased B-cell numbers and activation in CRSwNP patients compared to controls [59,60]. Moreover, Tertiary Lymphoid Organs (TLOs)[61], which harbour a defined B-cell compartment, have recently been described in sinonasal tissue [62] and reported to be more prevalent in CRSwNP compared to CRSsNP and controls [62,63].

We also found that 80.73% of the differentially-expressed genes in polyp tissue were downregulated (rather than upregulated) when compared to mucosa tissue, after controlling for sample pairing by patient. This "global reduced gene transcription state" in polyp tissue was also found in our GO analysis and is a novel finding. Whether this is due to a real widespread de-activation of transcription factors, or is merely an artifact of the relative paucity of cells in frank polyp tissue, is not certainly clear. Single-cell approaches (scRNA-Seq) could help address this question. Our finding mirrors previous conclusions by Platt et al. [64] and supports their recommendation that "care must be taken when collecting specimens for molecular studies of the sinonasal tract to differentiate polyp from nonpolyp tissue" [64]. Whether this recommendation has been followed consistently since then is not clear.

The CRSwNP_mucosa gene expression signature was more variable between samples and occupied an intermediate picture between the CRSsNP_mucosa, Control_mucosa and CRSwNP_polyp transcriptomic signatures. This might reflect disease in various stages of progression despite best efforts to gain a uniform tissue type for each group, in addition to known regional variation in innate immunity gene expression [65]. An alternative explanation is that in some patients with CRSwNP, the disease in its early stages could be exhibiting pathogenetic processes very similar to that found in CRSsNP disease. This variability of samples from different disease groups and the "intersection" of their transcriptomic signature on Principal Component Analaysis demonstrates the importance of CRS endotyping, and could be explained as highlighting the limitations of the current sNP/wNP paradigm in describing the underlying disease processes. On the other hand, the relatively strong signal seen originating from disease status in the transcriptomic signatures could be a marker that the classical CRS phenotypes still play an important role, not only in determining the broad clinical presentation of the patients, but also in determining the global transcriptomic signature and should not be also hastily discarded.

The small number of patients in each study group remains a limitation of our study, despite being (to the best of our knowledge) the largest bulk-tissue RNA-Seq study to date to be applied on CRS patients. Moreover, the patients included in this study were intentionally selected for their uniform phenotype (e.g. no asthma, no aspirin sensitivity) implying that more variation is to be expected if additional patients groups are to be recruited, adding further challenges to fully describing the broad heterogeneity of the disease. It also does not exclude a mixed endotype in the recruited cohort. Due to constraints of the selection criteria and absence of randomization, our cohort was dominated by male patients and is another limitation to be addressed in future studies with larger patient numbers. The effect of previous surgery is also yet to be explored. There remains a theoretical possibility that control patients with functional skullbase tumours can affect the sinus mucosa through a hormonal or paraneoplastic effect, but we are not aware of published evidence to that effect and they continue to be used as the standard control tissue in most CRS studies. With the progressive interest and advances in disease endotyping[3,4,66,67], further application of RNA-Seq-described transcriptomic signatures on patients of different endotypes would be an interesting avenue of future research, in order to validate the utility of this technique in deciphering the underlying pathology of the various CRS endotypes.

Conclusion

Our study represents one of the early (and to our knowledge, the largest to date) bulk-tissue RNA-Seq studies designed to explore the transcriptomic signatures in the sinonasal tissues of non-diseased control patients and CRS phenotypes and use these signatures to gain insights into the disease process. We report multiple findings that correlate with known clinical observations and our current understanding of the pathogenesis of chronic rhinosinusitis. Our results hold significant implication for future chronic rhinosinusitis research. We particularly stress the importance of tissue selection for molecular research utilizing human sinonasal tissue, especially in nasal polyposis, with evidence of a significant impact on the transcriptomic signature now considered well-established. The gene downregulation and the distinct signature found in polyp tissue may suggest that early intervention could be more effective in halting disease progression. The global transcriptomic signature showed some effect of disease phenotype, though highlighted the shortcoming of the sNP/wNP paradigm; but it remains unclear whether this techique would be sufficient to describe the complexity of the various CRS endotypes. This topic should be addressed in future research.

Acknowledgements

N/A.

Authorship contribution

AB conceived project idea and design; AB, JO sample collection, processing for RNA extraction and quality; AS, JG planned and designed sequencer setup; AB wrote primary and revised versions of manuscript, performed data analysis. AS, AT assisted with preliminary data analysis, provided critique and assistance with analysis. JG library preparation and sequencing. AJP, PJW, SV resources, senior supervision, manuscript critique. All authors read and approved the final manuscript.

Conflict of interest

Alkis J Psaltis is a consultant for Aerin Devices and ENT technologies and is on the speakers' bureau for Smith and Nephew, and received consultancy fees from Lyra Therapeutics. These are not relevant to this study.

Peter-John Wormald receives royalties from Medtronic, Integra, and Scopis, and is a consultant for NeilMed. These are not relevant to this study.

References

1. Fokkens WJ, Lund VJ, Mullol J, Bachert C, Alobid I, Baroody F, et al. EPOS 2012: European position paper on rhinosinusitis and nasal polyps 2012. A summary for otorhinolaryngologists. Rhinology. 2012;50:1–12.
2. Ponikau JU, Sherris DA, Kephart GM, Kern EB, Gaffey TA, Tarara JE, et al. Features of airway remodeling and eosinophilic inflammation in chronic rhinosinusitis: Is the histopathology similar to asthma? The Journal of Allergy and Clinical Immunology. 2003;112:877–82.
3. Bachert C, Akdis CA. Phenotypes and Emerging Endotypes of Chronic Rhinosinusitis. The Journal of Allergy and Clinical Immunology In Practice. 4:621–8.
4. Bachert C, Zhang N, Hellings PW, Bousquet J. Endotype-driven care pathways in patients with chronic rhinosinusitis. The Journal of Allergy and Clinical Immunology. 2018;141:1543–51.
5. Bassiouni A, Naidoo Y, Wormald P-J. When FESS fails: The inflammatory load hypothesis in refractory chronic rhinosinusitis. The Laryngoscope. 2012;122:460–6.
6. Liu Z, Kim J, Sypek JP, Wang I-M, Horton H, Oppenheim FG, et al. Gene expression profiles in human nasal polyp tissues studied by means of DNA microarray. The Journal of Allergy and Clinical Immunology. 2004;114:783–90.
7. Anand VK, Kacker A, Orjuela AF, Huang C, Manarey C, Xiang J. Inflammatory pathway gene expression in chronic rhinosinusitis. American Journal of Rhinology. 20:471–6.
8. Payne SC, Han JK, Huyett P, Negri J, Kropf EZ, Borish L, et al. Microarray analysis of distinct gene transcription profiles in non-eosinophilic chronic sinusitis with nasal polyps. American Journal of Rhinology. 22:568–81.
9. Sekigawa T, Tajima A, Hasegawa T, Hasegawa Y, Inoue H, Sano Y, et al. Gene-expression profiles in human nasal polyp tissues and identification of genetic susceptibility in aspirin-intolerant asthma. Clinical and Experimental Allergy. 2009;39:972–81.
10. Stankovic KM, Goldsztein H, Reh DD, Platt MP, Metson R. Gene expression profiling of nasal polyps associated with chronic sinusitis and aspirin-sensitive asthma. The Laryngoscope. 2008;118:881–9.
11. Rostkowska-Nadolska B, Kapral M, Fraczek M, Kowalczyk M, Gawron W, Mazurek U. A microarray study of gene expression profiles in nasal polyps. Auris, Nasus, Larynx. 2011;38:58–64.
12. Tyler MA, Russell CB, Smith DE, Rottman JB, Padro Dietz CJ, Hu X, et al. Large-scale gene expression profiling reveals distinct type 2 inflammatory patterns in chronic rhinosinusitis subtypes. The Journal of Allergy and Clinical Immunology. 2017;139:1061–1064.e4.
13. Okada N, Nakayama T, Asaka D, Inoue N, Tsurumoto T, Takaishi S, et al. Distinct gene expression profiles and regulation networks of nasal polyps in eosinophilic and non-eosinophilic chronic rhinosinusitis. International Forum of Allergy & Rhinology. 2018;8:592–604.
14. Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y. RNA-seq: An assessment of technical reproducibility and comparison with gene expression arrays. Genome Research. 2008;18:1509–17.
15. Wang Z, Gerstein M, Snyder M. RNA-Seq: A revolutionary tool for transcriptomics. Nature reviews Genetics. 2009;10:57–63.
16. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. Nature Methods. 2017;14:417–9.
17. Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, et al. GENCODE: The reference human genome annotation for The ENCODE Project. Genome Research. 2012;22:1760–74.
18. Frankish A, Diekhans M, Ferreira A-M, Johnson R, Jungreis I, Loveland J, et al. GENCODE reference annotation for the human and mouse genomes. Nucleic Acids Research. 2019;47:D766–73.
19. Soneson C, Love MI, Robinson MD. Differential analyses for RNA-seq: Transcript-level estimates improve gene-level inferences. F1000Research. 2016;4.
20. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology. 2014;15:550.
21. Ignatiadis N, Klaus B, Zaugg JB, Huber W. Data-driven hypothesis weighting increases detection power in genome-scale multiple testing. Nature Methods. 2016;
22. Van den Berge K, Soneson C, Robinson MD, Clement L. stageR: A general stage-wise

method for controlling the gene-level false discovery rate in differential expression and differential transcript usage. Genome Biology. 2017;18:151.
23. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium. Nature Genetics. 2000;25:25–9.
24. The Gene Ontology Consortium. Expansion of the Gene Ontology knowledgebase and resources. Nucleic Acids Research. 2017;45:D331–8.
25. Fabregat A, Jupe S, Matthews L, Sidiropoulos K, Gillespie M, Garapati P, et al. The Reactome Pathway Knowledgebase. Nucleic Acids Research. 2018;46:D649–55.
26. Young MD, Wakefield MJ, Smyth GK, Oshlack A. Gene ontology analysis for RNA-seq: Accounting for selection bias. Genome Biology. 2010;11:R14.
27. Hong G, Zhang W, Li H, Shen X, Guo Z. Separate enrichment analysis of pathways for up- and downregulated genes. Journal of the Royal Society, Interface. 2014;11:20130950.
28. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B (Methodological). 1995;57:289–300.
29. Lê Cao K-A, Boitard S, Besse P. Sparse PLS discriminant analysis: Biologically relevant feature selection and graphical displays for multiclass problems. BMC bioinformatics. 2011;12:253.
30. Rohart F, Gautier B, Singh A, Lê Cao K-A. mixOmics: An R package for 'omics feature selection and multiple data integration. PLoS computational biology. 2017;13:e1005752.
31. Kluyver T, Ragan-Kelley B, Pérez F, Granger B, Bussonnier M, Frederic J, et al. Jupyter Notebooks a publishing format for reproducible computational workflows. In: Loizides F, Scmidt B, editors. Positioning and Power in Academic Publishing: Players, Agents and Agendas. IOS Press; 2016. pp. 87–90.
32. Oliphant TE. Python for Scientific Computing. Computing in Science & Engineering. 2007;9:10–20.
33. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research. 2011;12:2825–30.
34. Anderson MJ. A new method for non-parametric multivariate analysis of variance. Austral Ecology. 2001;26:32–46.
35. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, et al. Vegan: Community Ecology Package. 2018.
36. Jackson DA. PROTEST: A PROcrustean Randomization TEST of community environment concordance. Écoscience. 1995;2:297–303.
37. Peres-Neto PR, Jackson DA. How well do multivariate data sets match? The advantages of a Procrustean superimposition approach over the Mantel test. Oecologia. 2001;129:169–78.
38. Wang W, Gao Z, Wang H, Li T, He W, Lv W, et al. Transcriptome Analysis Reveals Distinct Gene Expression Profiles in Eosinophilic and Noneosinophilic Chronic Rhinosinusitis with Nasal Polyps. Scientific Reports. 2016;6:26604.
39. Tokunaga T, Ninomiya T, Kato Y, Imoto Y, Sakashita M, Takabayashi T, et al. The significant expression of TRPV3 in nasal polyps of eosinophilic chronic rhinosinusitis. Allergology International: Official Journal of the Japanese Society of Allergology. 2017;66:610–6.
40. Ramakrishnan VR, Gonzalez JR, Cooper SE, Barham HP, Anderson CB, Larson ED, et al. RNA sequencing and pathway analysis identify tumor necrosis factor alpha driven small proline-rich protein dysregulation in chronic rhinosinusitis. American Journal of Rhinology & Allergy. 2017;31:283–8.
41. Cantero D, Cooksley C, Bassiouni A, Tran HB, Roscioli E, Wormald P-J, et al. Staphylococcus aureus biofilms induce apoptosis and expression of interferon- , interleukin-10, and interleukin-17A on human sinonasal explants. American Journal of Rhinology & Allergy. 2015;29:23–8.
42. Messerklinger W. Endoscopy of the nose. Urban & Schwarzenberg; 1978.
43. Stammberger H. Functional Endoscopic Sinus Surgery: The Messerklinger Technique. Decker; 1991.
44. Singh PK, Parsek MR, Greenberg EP, Welsh MJ. A component of innate immunity prevents bacterial biofilm development. Nature. 2002;417:552–5.
45. Rogan MP, Taggart CC, Greene CM, Murphy PG, O'Neill SJ, McElvaney NG. Loss of microbicidal activity and increased formation of biofilm due to decreased lactoferrin activity in patients with cystic fibrosis. The Journal of Infectious Diseases. 2004;190:1245–53.
46. Psaltis AJ, Bruhn MA, Ooi EH, Tan LW, Wormald P-J. Nasal mucosa expression of lactoferrin in patients with chronic rhinosinusitis. The Laryngoscope. 2007;117:2030–5.
47. Psaltis AJ, Wormald P-J, Ha KR, Tan LW. Reduced levels of lactoferrin in biofilm-associated chronic rhinosinusitis. The Laryngoscope. 2008;118:895–901.
48. Nadolska B, Frączek M, Kręcicki T, Kocięba M, Zimecki M. Lactoferrin inhibits the growth of nasal polyp fibroblasts. Pharmacological reports: PR. 62:1139–47.
49. Mollenhauer J, Wiemann S, Scheurlen W, Korn B, Hayashi Y, Wilgenbus KK, et al. DMBT1, a new member of the SRCR superfamily, on chromosome 10q25.3-26.1 is deleted in malignant brain tumours. Nature Genetics. 1997;17:32–9.
50. Rosenstiel P, Sina C, End C, Renner M, Lyer S, Till A, et al. Regulation of DMBT1 via NOD2 and TLR4 in intestinal epithelial cells modulates bacterial recognition and invasion. Journal of Immunology (Baltimore, Md: 1950). 2007;178:8203–11.
51. Lee RJ, Cohen NA. The emerging role of the bitter taste receptor T2R38 in upper respiratory infection and chronic rhinosinusitis. American Journal of Rhinology & Allergy. 27:283–6.
52. Kuntová B, Stopková R, Stopka P. Transcriptomic and Proteomic Profiling Revealed High Proportions of Odorant Binding and Antimicrobial Defense Proteins in Olfactory Tissues of the House Mouse. Frontiers in Genetics. 2018;9:26.
53. Browne J, Hopkins C, Slack R, van der Meulen J, Lund V, Topham J, et al. The national comparative audit of surgery for nasal polyposis and chronic rhinosinusitis. London: The Royal College of Surgeons of England; 2003.
54. Ko M-T, Chuang K-C, Su C-Y. Multiple analyses of factors related to intraoperative blood loss and the role of reverse Trendelenburg position in endoscopic sinus surgery. The Laryngoscope. 2008;118:1687–91.
55. Fraire ME, Sanchez-Vallecillo MV, Zernotti ME, Paoletti OA. Effect of premedication with systemic steroids on surgical field bleeding and visibility during nasosinusal endoscopic surgery. Acta Otorrinolaringologica Espanola. 64:133–9.
56. Grzegorzek T, Kolebacz B, Stryjewska-Makuch G, Kasperska-Zając A, Misiołek M. The influence of selected preoperative factors on the course of endoscopic surgery in patients with chronic rhinosinusitis. Advances in Clinical and Experimental Medicine: Official Organ Wroclaw Medical University. 23:69–78.
57. Mortuaire G, Bahij J, Maetz B, Chevalier D. Lund-Mackay score is predictive of bleeding in ethmoidectomy for nasal polyposis. Rhinology. 2008;46:285–8.
58. Asaka D, Nakayama T, Hama T, Okushi T, Matsuwaki Y, Yoshikawa M, et al. Risk factors for complications of endoscopic sinus surgery for chronic rhinosinusitis. American Journal of Rhinology & Allergy. 26:61–4.
59. Miljkovic D, Ou J, Kirana C, Hulse KE, Hauben E, Psaltis A, et al. Discordant frequencies of tissue-resident and circulating CD180-negative B cells in chronic rhinosinusitis. International Forum of Allergy & Rhinology. 2017;7:609–14.
60. Miljkovic D, Psaltis A, Wormald P-J, Vreugde S. Naive and effector B-cell subtypes are increased in chronic rhinosinusitis with polyps. American Journal of Rhinology & Allergy. 2018;32:3–6.
61. Neyt K, Perros F, GeurtsvanKessel CH, Hammad H, Lambrecht BN. Tertiary lymphoid organs in infection and autoimmunity. Trends in Immunology. 2012;33:297–305.
62. Lau A, Lester S, Moraitis S, Ou J, Psaltis AJ, McColl S, et al. Tertiary lymphoid organs in recalcitrant chronic rhinosinusitis. The Journal of Allergy and Clinical Immunology. 2017;139:1371–1373.e6.
63. Paramasivan S, Lester S, Lau A, Ou J, Psaltis AJ, Wormald P-J, et al. Tertiary lymphoid

organs: A novel target in patients with chronic rhinosinusitis. The Journal of Allergy and Clinical Immunology. 2018;142:1673–6.
64. Platt MP, Soler ZM, Kao S-Y, Metson R, Stankovic KM. Topographic gene expression in the sinonasal cavity of patients with chronic sinusitis with polyps. Otolaryngology–Head and Neck Surgery: 2011;145:171–5.
65. Seshadri S, Rosati M, Lin DC, Carter RG, Norton JE, Choi AW, et al. Regional differences in the expression of innate host defense molecules in sinonasal mucosa. The Journal of Allergy and Clinical Immunology. 2013;132:1227–1230.e5.
66. an BK, Klingler AI, Poposki JA, Stevens WW, Peters AT, Suh LA, et al. Heterogeneous inflammatory patterns in chronic rhinosinusitis without nasal polyps in Chicago, Illinois. The Journal of Allergy and Clinical Immunology. 2017;139:699–703.e7.
67. Stevens WW, Peters AT, Tan BK, Klingler AI, Poposki JA, Hulse KE, et al. Associations Between Inflammatory Endotypes and Clinical Presentations in Chronic Rhinosinusitis. J Allergy Clin Immunol Pract. 2019 May 22;

Professor Peter-John Wormald
3C: Department of Otolaryngology
Head and Neck Surgery
The Queen Elizabeth Hospital
28 Woodville Rd
Woodville South, SA 5011
Australia

Tel: +61 08 8222 7158
Fax: +61 08 8222 7419
E-mail:
peterj.wormald@adelaide.edu.au

This paper contains supplementary materials online: at www.rhinologyjournal.org