scientific reports

Check for updates

OPEN

Epigenome association study for DNA methylation biomarkers in buccal and monocyte cells for female rheumatoid arthritis

Gary Craig[1], Howard Kenney[1], Eric E. Nilsson[2], Ingrid Sadler-Riggleman[2], Daniel Beck[2] & Michael K. Skinner[2]✉

Genetics (i.e., mutations) has been assumed to be the major factor in rheumatoid arthritis (RA) etiology, but accounts for a minority of the variance in disease risk for RA. In contrast to genetics, the environment can have dramatic impacts on epigenetics that associate with disease etiology. The current study used buccal cells and purified blood monocytes from two different clinical cohorts involving Caucasian or African American female populations with or without arthritis. The differential DNA methylation regions (DMRs) between the control and RA populations were identified with an epigenome-wide association study. The DMRs (i.e., epimutations) identified in the buccal cells and monocytes were found to be distinct. The DMR associated genes were identified and many have previously been shown to be associated with arthritis. Observations demonstrate DNA methylation epimutation RA biomarkers are cell type specific and similar findings were observed with the two racial background populations. Rheumatoid arthritis susceptibility epigenetic diagnosis appears feasible and may improve the clinical management of RA and allowpreventative medicine considerations.

Abbreviations

RA	Rheumatoid arthritis
EWAS	Epigenome-wide association study
DMRs	Differential DNA methylation regions
GWAS	Genome-wide association study
MeDIP	Methylated DNA immunoprecipitation
MeDIP-Seq	Methylated DNA immunoprecipitation followed by next generation sequencing
LFC	Log-fold-change
PCA	Principal component analysis

Rheumatoid arthritis (RA) is a chronic autoimmune inflammatory disease predominant in females with a world-wide prevalence of 0.5–1.0%[1]. The frequency of RA has increased in the past few decades, and is proportionally higher in North American populations[1,2]. RA is characterized by synovial hyperplasia and joint destruction[3]. RA impacts both African American and Caucasian populations, but does have higher comorbidity[4] prevalence in African American populations. Coronary heart disease is more common in rheumatoid arthritis populations[4], as in other chronic inflammatory diseases. Inflammatory lung disease is a common extra-articular manifestation[5], along with rheumatoid nodules, secondary Sjogrens from endocrine gland inflammation, and more rarely rheumatoid vasculitis in this multisystem autoimmune disease. Correlations between RA and neurodegeneration have also been observed[6].

Environment and lifestyle have been shown to influence the risk of RA[7]. Smoking and nutrition are two of the major factors linked to RA risk[8,9]. Nutrition and dieting habits also have a role in RA risk and progression[10]. Environmental toxicant exposures have been shown to be involved in the etiology of rheumatoid arthritis[11,12]. In addition to environmental toxicants, alcohol consumption is a risk factor for the incidence of RA[13]. Similar observations have been shown in various worldwide populations and ethnic backgrounds[14–16]. Therefore, environmental factors and lifestyle have a significant impact on the etiology and progression of rheumatoid arthritis.

[1]Arthritis Northwest, 105 W 8th Ave #6080, Spokane, WA 99204, USA. [2]Center for Reproductive Biology, School of Biological Sciences, Washington State University, Pullman, WA 99164-4236, USA. ✉email: skinner@wsu.edu

Genetics has been assumed to be a major factor in rheumatoid arthritis etiology. Initially, gene associations were identified that involved a number of cellular pathways and immune related processes, such as the major histocompatibility complex (MHC), in particular the HLA-DRB1 and closely related genes[17]. These types of genetic mutation gene associations have been estimated to explain a minority of the variance in disease risk for RA[18,19]. A number of genome-wide association studies (GWAS) have been performed and identified hundreds of single-nucleotide polymorphisms (SNPs) that are associated with RA and speculated to impact a large number of biological processes[20]. Potential secondary gene associations suggest genetics can potentially explain 30% of familial disease cases[18]. Although these GWAS and similar gene impact studies have helped to better understand the molecular basis of RA, an alternate molecular process involving epigenetics is now assumed to be equally important and a significant factor in the etiology of RA[21]. Since environmental factors generally cannot directly change DNA sequence to alter genetic processes, the environmental impacts on RA etiology observed involve epigenetics. Epigenetics provides the molecular process for environmental factors such as nutrition and toxicants to impact genetics[22]. Therefore, an integration of environment, epigenetics and genetics is now thought to be involved in the etiology and progression of rheumatoid arthritis[21,23,24].

Epigenetics is defined as "molecular factors and processes around DNA that regulate genome activity independent of DNA sequence, and are mitotically stable"[22]. Epigenetic factors include DNA methylation, histone modifications, non-coding RNA (ncRNA), chromatin structure and RNA methylation[22]. Although all these processes will be involved in RA, DNA methylation at 5-methylcytosine has been the primary epigenetic process investigated in rheumatoid arthritis[21,25,26]. Epigenome-wide association studies (EWAS) have been used to identify specific immune gene associations[21,27], as have genome-wide investigations[21]. Some studies have investigated blood[28], which contains over 20 different cell types. In contrast to genetics, where the DNA sequence is the same between different cell types, each individual cell type has a unique epigenome to give the cell type its specificity. Therefore, mixed cell type (e.g., blood or mixed T cell lymphocytes) analysis can be misleading and reflect changes in cell populations instead of epigenetic change. A number of RA studies have investigated purified cell populations to provide insights into rheumatoid arthritis etiology including B lymphocytes[29], monocytes[30], and synoviocytes[31–34], which have distinct roles in RA etiology. Therefore, epigenetic analysis has provided insight into the pathology of rheumatoid arthritis[21,35].

The potential role of inheritance of RA has been demonstrated through familial clusters and parental transmission of arthritis susceptibility[36–38]. A non-genetic form of inheritance has been previously described that involves epigenetic alterations in the germline (sperm and egg) and the inheritance to subsequent generations, termed epigenetic transgenerational inheritance[22,39]. When the germ cell transmits the altered epigenetics to the developing embryo stem cells, all subsequent somatic cells developed will have altered epigenomes and transcriptomes. Those cell types sensitive to the shift in epigenetics will have a susceptibility to develop disease later in life[22]. Therefore, various somatic cell populations could be used as surrogate biomarker cell types to identify disease susceptibility and disease conditions. The role of epigenetic biomarkers in autoimmune disease and rheumatoid arthritis has previously been discussed[40,41]. A cell type population previously shown to have epigenetic alterations that is functionally related to disease activity in RA is the monocyte[30,42]. For the development of a biomarker cell type for disease susceptibility, the buccal cells are easily obtained from a cheek swab and have a high level of purity. Therefore, the current study used purified monocytes and buccal cells from a Caucasian population to develop a potential biomarker for rheumatoid arthritis. In addition, a distinct African American population cohort was also used with collection of buccal cells. This is one of the first observations that human buccal cells can be used as a surrogate marker cell for the detection of disease associated epimutations. Both Caucasian and African American population clinical cohorts were used to compare non-arthritis individuals versus arthritis individuals to identify RA associated DNA methylation alterations, termed epimutations. A combined analysis of Caucasian and African American populations provided an epigenetic biomarker in buccal cells for RA that was more efficient for diagnosis of RA susceptibility than either population alone. These epimutations can potentially be used as a biomarker for RA to improve the detection and clinical management[43,44] of the disease. The current study provides the proof of concept that buccal cell RA epigenetic biomarkers may exist, however, future larger clinical trial studies are required to optimize the epigenetic biomarker.

Results

Females of similar age and race were collected for comparison of individuals with and without rheumatoid arthritis, Table 1. The white Caucasian non-Hispanic female samples were obtained by the Arthritis Northwest (ANW) Clinic in Spokane, Washington. The African American (AA) female samples were obtained by Dx Biosamples, LLC in San Diego, California, with sample collections in the Los Angeles California area. Institutional review board (IRB) approvals were obtained for the study from both sources. The individual sample information for age, race, and clinical information is presented in Table 1. Upon collection of buccal cell swabs, the samples were frozen at − 20 °C then shipped on dry ice and stored at − 80 °C until use. The procedure and sample processing for buccal cells is presented in the Supplemental Methods. For Caucasian ANW clinic patients, a blood sample was also obtained and shipped immediately on ice for isolation of monocytes, as described in the Methods. Monocytes were purified with an antibody bead procedure, as previously described[45]. After isolation, the monocytes were stored at- − 80 °C prior to use. The Caucasian (CC) buccal cell control without RA (n = 13) and with RA (n = 13) were compared. The Caucasian monocyte cell control without RA (n = 13) and with RA (n = 13) were compared. The African American (AA) buccal cell control without RA (n = 9) and with RA (n = 13) were compared. A combination of Caucasian and African American samples without (n = 23) and with RA (n = 26) were compared. The individual sample information is presented in Table 1 and provides clinical information and RA diagnostic information. The Caucasian samples had RA diagnostic assays of Rheumatoid Factor autoantibody

Subject #	Group	Age	Race	Gender	RA Diagnostic Information
(A) Arthritis Northwest Caucasian Population					
ANWC01	Control	74	Caucasian	F	Not Applicable
ANWC02	Control	56	Caucasian	F	Not Applicable
ANWC03	Control	39	Not specified	F	Not Applicable
ANWC04	Control	56	Caucasian	F	Not Applicable
ANWC05	Control	59	Caucasian	F	Not Applicable
ANWC06	Control	68	Caucasian	F	Not Applicable
ANWC07	Control	55	Caucasian	F	Not Applicable
ANWC08	Control	70	Caucasian	F	Not Applicable
ANWC09	Control	49	Caucasian	F	Not Applicable
ANWC10	Control	62	Caucasian	F	Not Applicable
ANWC11	Control	30	Caucasian	F	Not Applicable
ANWC12	Control	51	Caucasian	F	Not Applicable
ANWC13	Control	45	Caucasian	F	Not Applicable
	Mean age ± SEM	54.9 ± 3.4			
ANWT01	Rheumatoid arthritis	73	Caucasian	F	High RA activity, RF +, CCP +, CDAI +
ANWT03	Rheumatoid arthritis	61	Caucasian	F	High RA activity, RF +, CCP +, CDAI +
ANWT04	Rheumatoid arthritis	41	Caucasian	F	High RA activity, RF +, CCP +, CDAI +
ANWT05	Rheumatoid arthritis	56	Caucasian	F	High RA activity, RF +, CCP +, CDAI +
ANWT06	Rheumatoid arthritis	57	Caucasian	F	High RA activity, RF +, CCP +, CDAI +
ANWT09	Rheumatoid arthritis	66	Caucasian	F	High RA activity, RF +, CCP +, CDAI +
ANWT10	Rheumatoid arthritis	50	Caucasian	F	Moderate RA activity, RF +, CDAI +
ANWT12	Rheumatoid arthritis	74	Caucasian	F	Moderate RA, activity RF +, CDAI +
ANWT13	Rheumatoid arthritis	54	Caucasian	F	High RA activity, RF +, CCP +, CDAI +
ANWT14	Rheumatoid arthritis	68	Caucasian	F	High RA activity, RF +, CCP +, CDAI +
ANWT15	Rheumatoid arthritis	26	Caucasian	F	High RA activity, RF +, CCP +, CDAI +
ANWT16	Rheumatoid arthritis	48	Caucasian	F	High RA activity, RF +, CCP +, CDAI +
ANWT17	Rheumatoid arthritis	48	Native American	F	High RA activity, RF +, CCP +, CDAI +
	Mean age ± SEM	55.5 ± 3.7			
(B) DxBiosamples African American Population					
AH_10	Control	55	African American	F	Not Applicable
AH_12	Control	52	African American	F	Not Applicable
AH_13	Control	55	African American	F	Not Applicable
AH_14	Control	57	African American	F	Not Applicable
AH_15	Control	52	African American	F	Not Applicable
AH_16	Control	58	African American	F	Not Applicable
AH_17	Control	65	African American	F	Not Applicable
AH_18	Control	61	African American	F	Not Applicable
AH_19	Control	57	African American	F	Not Applicable
	Mean age ± SEM	56.8 ± 1.3			
AH_20	Rheumatoid arthritis	55	African American	F	High RA activity, CDAI +
AH_22	Rheumatoid arthritis	51	African American	F	High RA activity, CDAI +
AH_23	Rheumatoid arthritis	53	African American	F	High RA activity, CDAI +
AH_24	Rheumatoid arthritis	56	African American	F	Moderate RA activity, CDAI +
AH_25	Rheumatoid arthritis	58	African American	F	High RA activity, CDAI +
AH_26	Rheumatoid arthritis	52	African American	F	High RA activity, CDAI +
AH_27	Rheumatoid arthritis	53	African American	F	Moderate RA activity, CDAI +
AH_28	Rheumatoid arthritis	50	African American	F	High RA activity, CDAI +
AH_29	Rheumatoid arthritis	51	African American	F	High RA activity, CDAI +
AH_30	Rheumatoid arthritis	50	African American	F	High RA activity, CDAI +
AH_31	Rheumatoid arthritis	52	African American	F	High RA activity, CDAI +
AH_32	Rheumatoid arthritis	52	African American	F	High RA activity, CDAI +
AH_33	Rheumatoid arthritis	63	African American	F	High RA activity, CDAI +
	Mean age ± SEM	53.5 ± 1.0			

Table 1. Buccal and monocyte samples and clinical information. The samples were provided by ANW Clinic, Spokane, WA and Dx Biosamples, San Diego, CA, that collected and obtained the clinical information of age, sex, clinical diagnosis RA and site collection. The mean ± SD for case and control samples are provided for age. A student t-test was used to assess any statistical differences between RA and control groups, and none were found. *RF +* Rheumatoid Factor autoantibody positive, *CCP +* Cyclic citrullinated peptide autoantibody positive, *CDAI* Clinical Disease Activity Index positive with high RA activity > 22 and moderate activity 10–22. CDAI supportive assays at DAS28 and RAPID3 positive.

(RF) and Citrullinated Peptide autoantibody (CCP), and more qualitative RA activity assays of Clinical Disease Activity Index (CDAI) with DAS28 and RAPID3 supportive analysis, Table 1.

The DNA was extracted from each sample as described in the Supplemental Methods. The DNA was then sonicated to 150–300 bp fragments and used for a methylated DNA immunoprecipitation (MeDIP) protocol, as described in the Supplemental Methods. This involved an antibody to 5-methylcytosine and a magnetic bead procedure[46]. This immunoprecipitated methylated DNA was then used to generate a sequencing library for an MeDIP-Seq analysis, as described in the Supplemental Methods. MeDIP-Seq allows for greater than 90% of the genome to be examined for this EWAS analysis. Each sample for the MeDIP-Seq analysis had approximately 25 million reads, and the quality control details are presented in the Supplemental Methods. The EdgeR statistical analysis was used to identify differential DNA methylation regions (DMRs). Various *p*-value thresholds are presented for each of the comparisons, and $p < 1e-04$ was selected for subsequent analysis, Fig. 1. The majority of DMRs had one significant 1 kb window, but some had multiple windows. The Caucasian (CC) control versus RA buccal had 362 DMRs (Fig. 1a), CC control versus RA monocytes 617 DMRs (Fig. 1b), AA control versus AA RA buccal had 364 DMRs (Fig. 1c), and combination of CC and AA RA all buccal had 308 DMRs (Fig. 1d). A venn diagram of the different comparison DMRs at $p < 1e-04$ had no overlap, except for the all buccal combined CC and AA that had approximately a 10% overlap, Fig. 1e. An extended overlap of the $p < 1e-04$ DMR comparisons with the others at a reduced statistical threshold of $p < 0.05$ was used to determine if a DMR overlap is present at a reduced threshold, Fig. 1f. From the horizontal row, the same cell type has 100% overlap, as anticipated. The same row allows potential overlaps at a reduced statistical threshold to be identified. The overlap was 11% between the Caucasian monocyte and buccal cell. The other overlaps were between 3 and 7%, Fig. 1f. Interestingly, the combined CC and AA (all buccal) analysis had a 74% overlap with the CC buccal and 87% with the AA buccal (Fig. 1f highlight), suggesting the combined analysis DMR set is optimal for an RA biomarker. Therefore, the majority of the DMRs for each specific distinct comparison were cell and race specific, while the combined CC and AA analysis had good overlap, Fig. 1f.

The information for each comparison set of DMRs is presented in Supplemental Tables S1, S2, S3 and S4. The DMR name, chromosomal location, start and stop nucleotide sites, length bp, number of 1 kb significant windows, EdgeR *p*-value, maximum log-fold change (maxLFC) (i.e., positive values being an increase in DNA methylation, and negative values a decrease in DNA methylation), CpG number and density, and DMR associated gene and gene category, are presented in Supplemental Tables S1–S4. The chromosomal locations of the DMRs are presented in Fig. 2a–d with chromosome number and location (megabase) indicated for each with a red arrowhead. The black boxes indicate a cluster of DMRs. All the chromosomes contain DMRs with this genome-wide analysis, Fig. 2a–d. Therefore, the RA DMRs (e.g., epimutations) are genome-wide. The DMR characteristics demonstrate a low CpG density, termed a CpG desert, of 1–3 CpG/100 bp, Supplemental Figure S1, for all the comparison DMRs. The sizes of the DMRs are predominantly 1 kb in size with some being 2–4 kb in size, Supplemental Figure S1, for all the comparisons. A comparison of the principal components in a Principal Component Analysis (PCA) demonstrates that the DMR components for the control versus RA cluster separately for each of the comparison groups, Fig. 3a–d. The only exception is with the Caucasian control versus RA monocytes, Fig. 3b, that has one control and one RA each that overlap with the PCA cluster. The overlapping RA group DMR at PC1 = − 8 and PC2 = 2.2 was 74 year age. The other individuals at > 70 year age were the DMRs at PC1 = − 4 and PCR = − 18, PC1 = 31 and PC2 = − 0.5, and PC1 = − 4 and PC2 = 9.2, Fig. 3b. Higher variability was observed with the > 70 year individuals. Therefore, the PCA indicates principal components of the DMRs are predominantly distinct between the control and arthritis groups. In addition, the Native American and individual not specified for race (Table 1) were not outliers with the PCA analysis (Fig. 3), so we included in the analysis. The Caucasian sample sets had a few 70 year and 30 year outliers compared to the mean of approximately 55 year samples, Table 1a. Reanalysis of the comparison sample sets deleting these samples did not impact significantly the DMR number nor statistics. Future studies are needed with larger sample sizes to more accurately determine the impact of age on the RA epigenetic biomarkers. The genome-wide statistical analysis with EdgeR demonstrated the monocyte analysis provided a false discovery rate (FDR) of < 0.1, but the buccal analysis gave an FDR of approximately 0.2. Therefore, the buccal cell DMR analysis was more variable than the monocyte analysis. The significance of each DMR is indicated with minimum *p*-value and minimum FDR in the Supplemental Tables S1–S4. Future studies require increased sample size when using buccal cells as a marker cell. The potential DMR biomarkers for rheumatoid arthritis identified demonstrated predominantly unique chromosomal locations for each comparison (Fig. 2), but similar genomic features, Supplemental Tables S1–S4. The RA DMR biomarkers appear cell type specific, and are distinct between the Caucasian and African American populations when analyzed separately. However, a combined CC and AA all buccal cell analysis did show strong overlap among the groups, Fig. 1f.
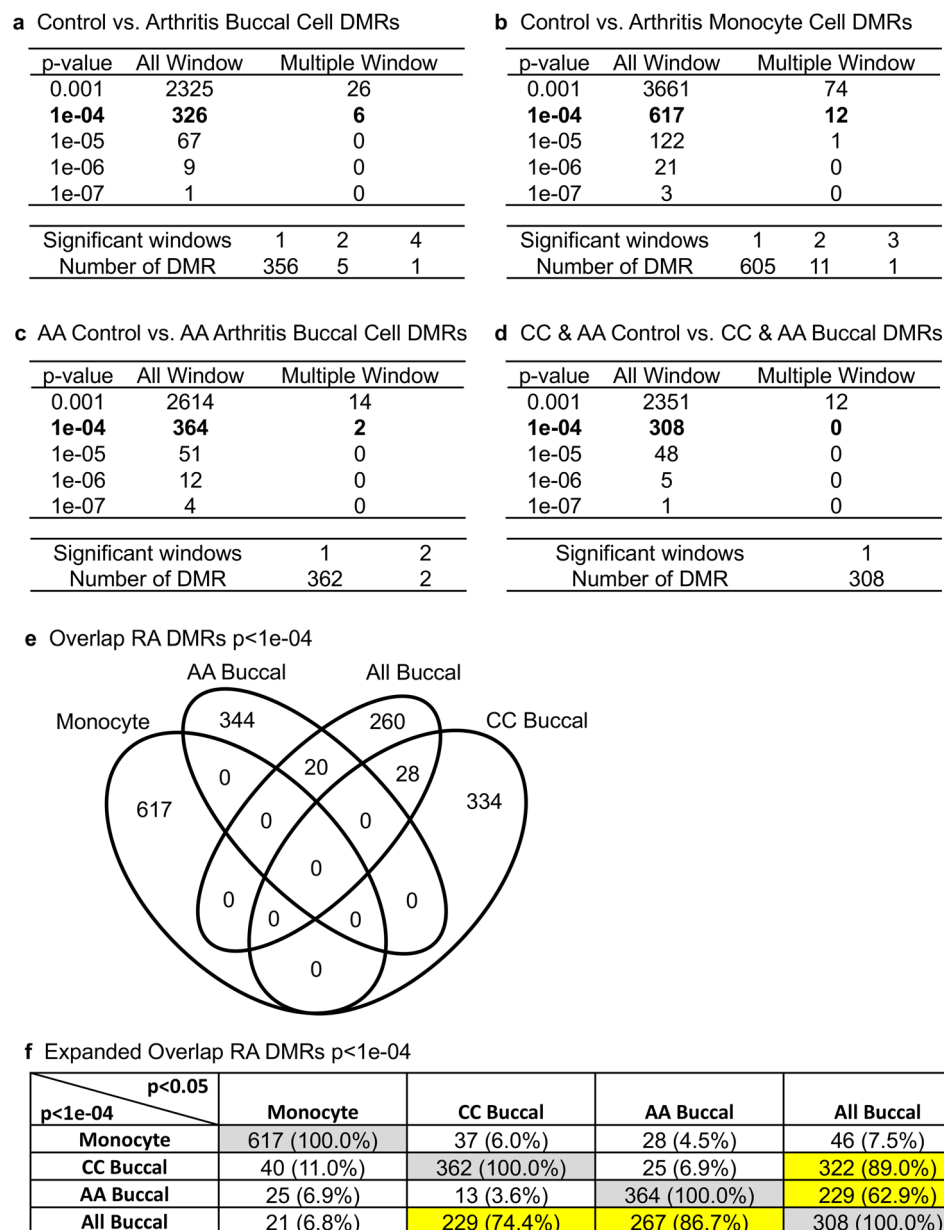
a Control vs. Arthritis Buccal Cell DMRs

p-value	All Window	Multiple Window		
0.001	2325	26		
1e-04	**326**	**6**		
1e-05	67	0		
1e-06	9	0		
1e-07	1	0		
Significant windows	1	2	4	
Number of DMR	356	5	1	

b Control vs. Arthritis Monocyte Cell DMRs

p-value	All Window	Multiple Window		
0.001	3661	74		
1e-04	**617**	**12**		
1e-05	122	1		
1e-06	21	0		
1e-07	3	0		
Significant windows	1	2	3	
Number of DMR	605	11	1	

c AA Control vs. AA Arthritis Buccal Cell DMRs

p-value	All Window	Multiple Window	
0.001	2614	14	
1e-04	**364**	**2**	
1e-05	51	0	
1e-06	12	0	
1e-07	4	0	
Significant windows	1	2	
Number of DMR	362	2	

d CC & AA Control vs. CC & AA Buccal DMRs

p-value	All Window	Multiple Window
0.001	2351	12
1e-04	**308**	**0**
1e-05	48	0
1e-06	5	0
1e-07	1	0
Significant windows		1
Number of DMR		308

e Overlap RA DMRs p<1e-04

f Expanded Overlap RA DMRs p<1e-04

p<1e-04 \ p<0.05	Monocyte	CC Buccal	AA Buccal	All Buccal
Monocyte	617 (100.0%)	37 (6.0%)	28 (4.5%)	46 (7.5%)
CC Buccal	40 (11.0%)	362 (100.0%)	25 (6.9%)	322 (89.0%)
AA Buccal	25 (6.9%)	13 (3.6%)	364 (100.0%)	229 (62.9%)
All Buccal	21 (6.8%)	229 (74.4%)	267 (86.7%)	308 (100.0%)

Figure 1. Rheumatoid arthritis (RA) DMR identifications. **(a)** Caucasian control versus RA buccal cell DMR analysis. **(b)** Caucasian control versus RA monocyte cell DMR analysis. **(c)** African American (AA) control versus RA buccal cell DMR analysis. **(d)** Combined Caucasian (CC) and African American (AA) control versus RA buccal cell DMR analysis. The number of DMRs found using different p-value cutoff thresholds. The all window column shows all DMRs. The multiple window column shows the number of DMRs containing at least two adjacent significant windows and the number of DMRs with each specific number of significant windows at a p-value threshold of $p < 1e-04$. **(e)** Venn diagram overlap of the RA DMRs at $p < 1e-04$ for the Caucasian monocyte, buccal and AA buccal, and combined CC and AA (All) buccal. **(f)** Extended overlap with a comparison of RA DMRs in the different comparison at $p < 1e-04$ versus horizontal $p < 0.05$ for the different comparisons. The overlapping DMR numbers and percent (%) of the total is presented. The highlighted overlaps for All buccal and overlaps indicated.

Less than half of the RA DMRs identified with each comparison had DMR associated genes within 10 kb of the DMR, Supplemental Tables S1–S4. The 10 kb distance is used to include the proximal and distal promoter regions of the genes. The RA DMR associated gene categories were identified and are presented in Fig. 4a. The DMR numbers for each associated gene functional category are presented for each comparison. The signaling, transcription, metabolism and receptor are predominant for each of the DMR comparisons. This does reflect the major gene categories within the human genome with metabolism, transcription and signaling being the largest gene categories. This gene category analysis was for an individual DMR gene association analysis and
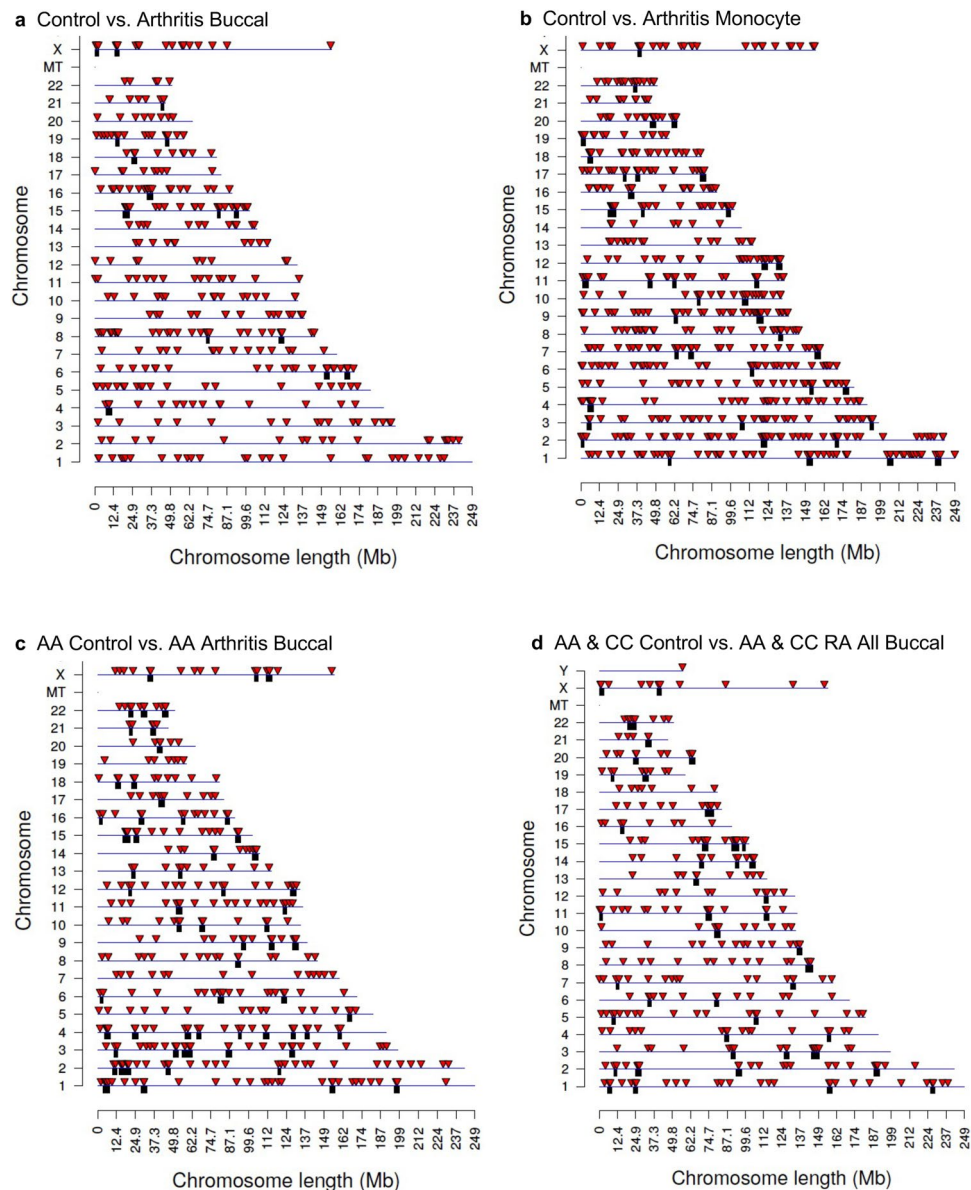
Figure 2. RA DMR chromosomal locations and principal component analysis (PCA). The DMR locations on the individual chromosomes are identified. All DMRs at a *p*-value threshold of $p < 1e-04$ are shown with the red arrowheads and clusters of DMRs with the black boxes. **(a)** Caucasian control versus RA buccal DMRs. **(b)** Caucasian control versus RA monocyte DMRs. **(c)** AA control versus RA buccal DMRs. **(d)** Combined CC and AA for All buccal control versus RA buccal DMRs. All DMRs at a *p*-value threshold of $p < 1e-04$.

not to reflect group combined function. In contrast, a gene pathway analysis (KEGG) was performed, and those pathways in two or more different comparison DMR associated genes with similar function were identified in Fig. 4b. The metabolic pathway was excluded as it is present in all comparisons and over-represented in pathway analysis. The common signaling pathways include pathways in cancer, pathways in neurodegenerative disease, and specific pathways such as P13K-Akt, Fig. 4b.

The final analysis correlated the RA DMR associated genes with pathologies and processes. Pathway Studio was used to identify the associated pathologies, as described in the Supplemental Methods. The African American RA buccal cell DMR associated gene correlations to diseases are presented in Fig. 5a. Six of the major disease correlations were arthritis related. The others are cancer and neurodegenerative related. Therefore, the predominant African American RA DMRs in buccal cells are associated with arthritis pathologies, Fig. 5a. The Caucasian buccal cell RA DMR associated gene correlations with rheumatoid arthritis and arthritis pathologies are shown, Fig. 5b, with contiguous gene syndrome and intellectual disability as additional pathologies significantly correlated. The combination CC and AA analysis also provided DMR associated genes and arthritis gene associations, Fig. 6a. The Caucasian monocyte DMRs also predominantly associated with arthritis and RA correlated genes, with additional epilepsy and carcinoma associated genes, Fig. 6b. Observations demonstrate
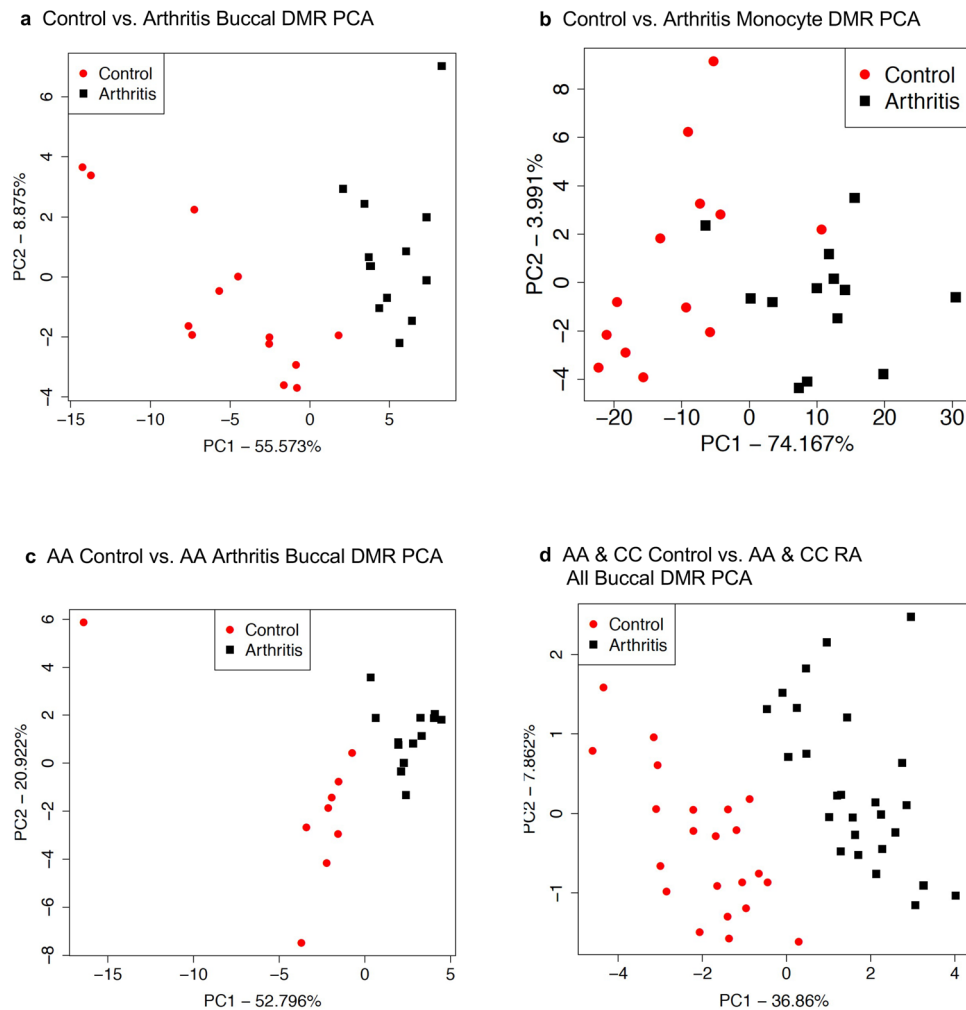
Figure 3. Control versus RA DMR principal component analysis (PCA). PCA analysis for DMRs at $p < 1e-04$. **(a)** Caucasian control versus RA. **(b)** Caucasian monocyte DMR PCA. **(c)** AA control versus AA RA buccal DMR PCA. **(d)** CC and AA all buccal control versus all RA buccal DMR PCA.

the RA DMRs identified with all the comparisons showed significant connections to genes correlated to arthritis pathologies, including rheumatoid arthritis. An additional analysis used the DMR associated genes for RA from each of the comparisons to identify known RA cellular process correlations, Fig. 7. Previous rodent and human studies have identified known gene correlations with RA associated cell processes, and these same genes and correlations were observed in the current study, Fig. 7 and Supplemental Table S5. Over 25 different RA gene cell processes were identified with a statistical significance $p < 1e-06$, with the four most significant being protein regulators of immune response, inflammatory response, innate immune response, and cellular immune response, Supplemental Table S5. These DMR associated rheumatoid arthritis gene cell process correlations are shown in Fig. 7 and Supplemental Table S5. Therefore, the RA DMR associated genes identified in the current study correlated to previously identified RA associated genes and cellular processes, as well as identified potential new RA associated genes.

Discussion

Rheumatoid arthritis impacts approximately one percent of the worldwide population, and some areas such as North America have a higher incidence of RA among the population[21]. RA is generally diagnosed today after the onset of the disease when clinical characteristics develop. Therefore, most diagnosis is based on symptomatic parameters. Although several RA biomarkers and diagnostics have been developed[47,48], none are focused on early-stage biomarkers that can be used prior to the onset of clinical symptoms. The advantage of epimutations such as DNA methylation biomarkers is that they have been shown to develop early in life, or through epigenetic inheritance ancestrally, such that the biomarkers have the potential to act as diagnostics for preventative medicine treatments[41,49]. Such an early-stage RA biomarker or diagnostic has not been developed and is the focus of the current study. The clinical management and treatment of the disease has been shown to be more effective as a preventative treatment prior to disease onset. An example is the use of the chemotherapy tamoxifen, which has

a RA DMR Associated Gene Categories p<1e-04
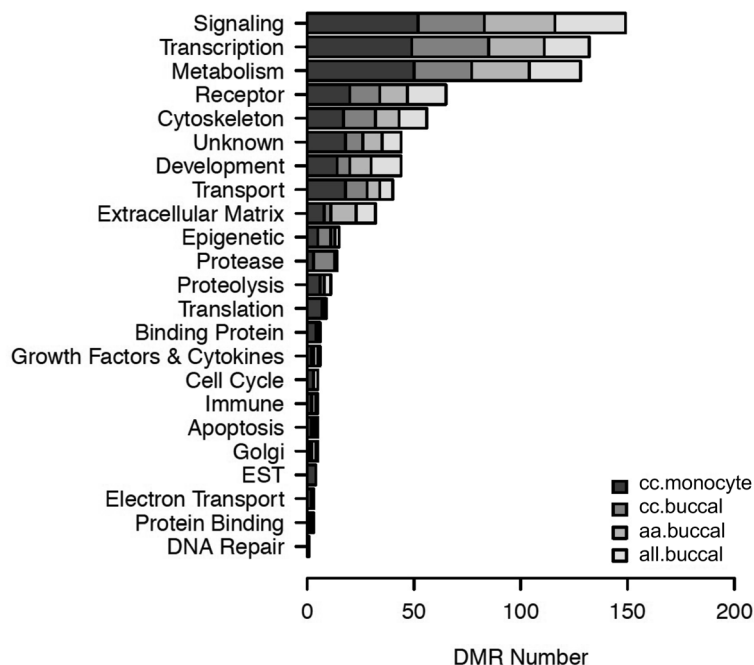
b RA DMR Associated Gene Pathways (KEGG) p<1e-04

Arthritis RA CC Buccal 1e-04
hsa05016 Huntington disease (8)
hsa04360 Axon guidance (7)
hsa05022 Pathways of neurodegeneration (7)
hsa04144 Endocytosis (7)
hsa05200 Pathways in cancer (6)

Arthritis RA CC Monocyte
hsa05022 Pathways of neurodegeneration (12)
hsa05200 Pathways in cancer (12)
hsa04151 PI3K-Akt signaling pathway (12)
hsa04010 MAPK signaling pathway (8)
hsa04144 Endocytosis (8)

Rheumatoid Arthritis RA AA Buccal 1e-04
hsa01100 Metabolic pathways (21)
hsa05165 Human papillomavirus infection (9)
hsa05200 Pathways in cancer (9)
hsa04151 PI3K-Akt signaling pathway (8)
hsa04360 Axon guidance (7)

Rheumatoid Arthritis RA All Buccal
hsa01100 Metabolic pathways (17)
hsa05200 Pathways in cancer (10)
hsa05022 Pathways of neurodegeneration (10)
hsa04723 Retrograde endocannabinoid signaling (7)
hsa04024 cAMP signaling pathway (7)

Figure 4. RA DMR associated gene categories and pathways. **(a)** DMR associated gene categories. DMR numbers at a *p*-value threshold *p* < 1e−04 are shown. The comparison DMR key is inset. **(b)** DMR associated gene pathways. The pathways common for two or more comparisons are presented. Number in bracket is number of DMR associated genes in pathway.

a low efficiency in treating breast cancer once the cancer has developed. However, tamoxifen can be effective as a preventative treatment for a female in her 30 s prior to disease development[50]. The possibility that current RA therapeutics may be more efficient as preventative treatments prior to the onset of disease remains to be investigated. The key factor in allowing these types of preventative therapeutics to be considered is the availability of an RA biomarker that can be used prior to the onset of disease or symptoms. The current study is designed to develop an epigenetic biomarker for RA that, due to the nature of environmental impacts on developmental programming of epigenetics, can potentially also act as an RA susceptibility diagnostic to facilitate a preventative medicine approach for the disease.

Genome-wide association studies (GWAS) that are used to identify genetic mutations have generally found that less than 1% of the population with a specific disease has correlated genetic mutations. This has become the major factor inhibiting effective genetic diagnostics for disease[51]. In contrast, epigenome-wide association studies (EWAS) often identify epimutations that are present in the majority of the population with a specific disease[52]. This allows EWAS studies to use smaller sample sizes compared to GWAS studies. The current study found RA DNA methylation DMR sites were present in the populations with RA when compared to those without, Fig. 1a–c. As previously demonstrated, epigenetic biomarkers are anticipated to be more efficient as a tool for disease and disease susceptibility analysis[53,54]. In the current study, two distinct control and RA clinical cohort populations were obtained for analysis. The Caucasian and African American populations were obtained at different locations under separate IRBs. Similar observations were made with both these clinical cohorts. Although the current study has the limitation of not having a blinded test set or distinct expanded clinical cohort, the current study does use two distinct clinical populations from different racial backgrounds, locations, and clinical sites for comparison. Interestingly, the Caucasian and African American population buccal cells had

a Arthritis AA Buccal DMR Associated Gene Disease Correlations
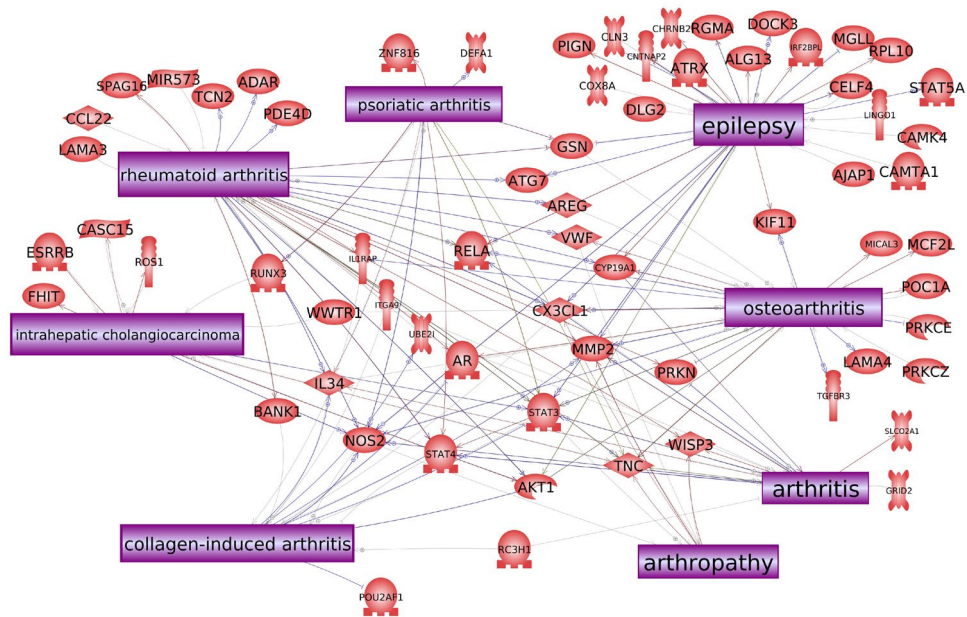
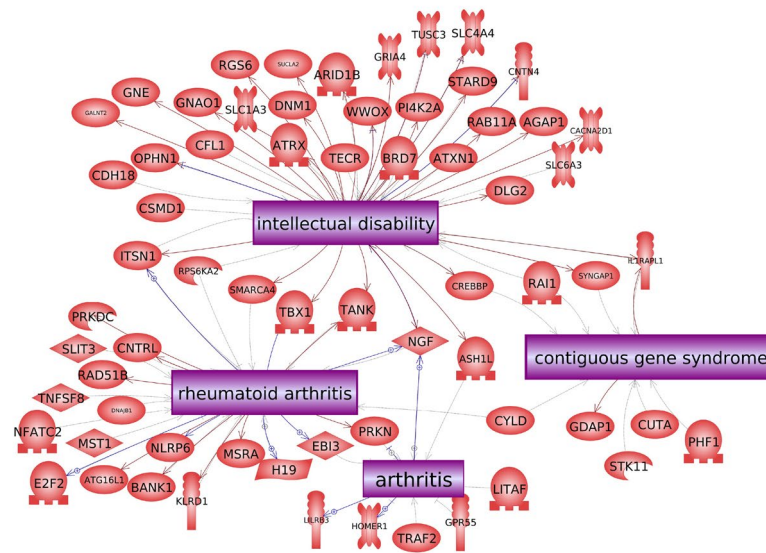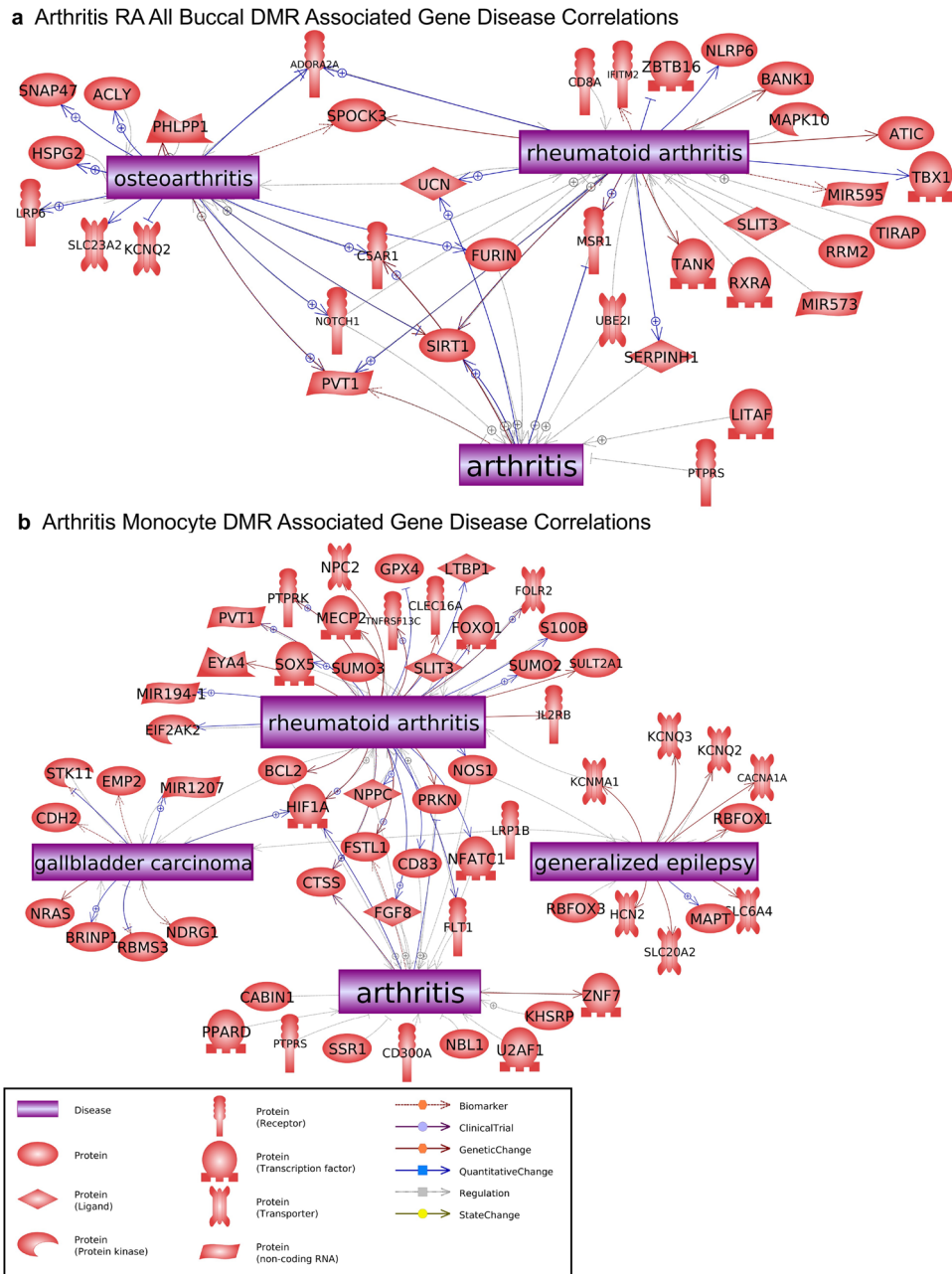**b** Arthritis CC Buccal DMR Associated Gene Disease Correlations

Figure 5. DMR associated genes from the current study were compared to genes associated with arthritis in the published literature using Pathway Studio software (Elsevier, Inc.). Those that were in common are depicted. **(a)** African American (AA) buccal cell RA DMR associated gene disease correlations. **(b)** Caucasian buccal DMR associated gene disease correlations.

distinct RA DMRs with only a 3.6% overlap of 13 DMRs when analyzed separately, Fig. 1e. Observations suggest that ethnic background and race need to be considered in the development of epigenetic DNA methylation diagnostics for disease. One variable in this analysis is that the Caucasian samples were collected in Spokane, Washington, while the African American samples were collected in Los Angeles, California. Since epigenetics is environmentally responsive and inherited[39], the demographics and collection regions will need to be considered. Spokane Washington is a moderate size city with less environmental contaminants versus Los Angeles, which is a large urban city with higher levels of pollutant exposure such as air pollution. The generational backgrounds will also be distinct between the populations, and ethnic and racial impacts on epigenetics has been observed[55]. However, when the CC and AA populations were combined, a DMR set with strong overlap with both the individual comparisons was identified, Fig. 1f. Therefore, expanded future studies should incorporate diverse race and ethnic background in the development of epigenetic DMR rheumatoid arthritis diagnostics. Observations from the current study demonstrate similar levels of epigenetic change (i.e., numbers of RA DMRs observed), but the majority of DMRs were distinct between the Caucasian and African American populations. This is one of the first observations for potential race and ethnic background impacts on the development of epigenetic

Figure 6. DMR associated genes from the current study were compared to genes associated with arthritis in the published literature using Pathway Studio software (Elsevier, Inc.). Those that were in common are depicted. **(a)** Combined CC and AA all buccal RA DMR associated gene disease correlations. **(b)** Caucasian monocyte RA DMR associated gene disease correlations. The gene function symbol index inset.

disease biomarkers. Interestingly, the combination of the populations within the analysis generated a more efficient epigenetic biomarker for female RA susceptibility. Although the current study only provides the proof of concept potential biomarkers for RA may be developed, expanded clinical trials with larger sample size are now required. Diverse populations need to be considered in future studies and potential commercialization of such disease biomarkers.

The current study in the Caucasian population compared two different cell populations for the analysis of potential RA biomarkers. The monocytes are an immune related cell that has been shown to develop molecular alterations in RA patients relevant to the disease pathology[56]. As previously described, the alterations in epigenetics observed in the current study are relevant to RA, Figs. 6 and 7. In contrast, the buccal cells from a cheek swab have no distinct relationship to the RA disease pathology or etiology, but functions as a surrogate marker cell for epigenetic alterations. Buccal cells are one of the easiest purified cell populations to collect, which is required for epigenetics, and they are the least invasive for the patient. The blood collection, followed by monocyte cell

Figure 7. DMR associated RA genes were correlated with known RA cell processes in the published literature using Pathway Studio software (Elsevier, Inc.). The DMR associated genes from all comparisons were linked to the RA cell processes.

purification, would not be an efficient biomarker cell to develop for RA compared to the buccal cell. Since the majority of disease has an early developmental exposure or impact that promotes the developmental origins of health and disease (DOHaD), all cell types in the body can be impacted by these early life events. Epigenetics is the primary mechanism involved in these developmental origins of disease, since genetic mutations are a minor component of disease correlations. Therefore, ancestral and early life exposure can impact the epigenetics of all cell types in the body such that marker cells like the buccal cells can be used to reflect epigenetic alterations associated with later life disease susceptibility and diagnosis. Although the buccal cells are simply a marker surrogate cell for epigenetic alterations, the current study demonstrates that a number of the RA DMRs identified do have associations with RA and generational arthritis, Figs. 5, 6 and 7. One of the major observations of the current study is that easily collected buccal cells may act as an effective surrogate marker cell for the epigenetic diagnosis of disease susceptibility.

Many of the DMR associated genes correlated with previously identified RA disease associated genes, Figs. 5, 6 and 7. The DMR associated genes were within 10 kb of the gene to consider the proximal and distal promoters of the genes. Although some DMRs are only within the promoter region, the majority of DMRs overlapped with the gene. The RA DMR associated genes suggest a potential regulatory role, but future studies are needed to demonstrate a potential to regulate gene expression. Approximately 50% of the DMRs have an increase in DNA methylation and the rest a decrease in DNA methylation, Supplemental Tables S1–S4 (log fold change, max LFC value). Interestingly, the DMR associated genes for the Caucasian buccal cell and monocyte cells both had RA associated genes, but were distinct, Figs. 5, 6 and 7. Some of the RA associated genes common among the different comparisons included Sox5, Fox01, FGF8, TNFSF8, TANK and H19. Two genes not in the DMR associated genes at $p < 1e-04$ that have been shown to be correlated with RA are HLA-DRB1[57] and 14-3-3eta[58]. The HLA-DRB1 is within the immune complex gene group and is associated with RA therapeutics[59]. The 14-3-3 eta gene has been shown to be a useful RA biomarker[60]. Further analysis at a reduced statistical threshold demonstrated DMRs within 10 kb of these genes to include the proximal and distal promoter region for HLA-DRB1 in the African American buccal ($p < 0.02$), and the 14-3-3 eta in the Caucasian monocyte ($p < 0.0003$) and buccal ($p < 0.03$), and African American buccal ($p < 0.04$). Therefore, the DMRs at reduced statistical thresholds will also be important to consider in the RA disease etiology. Previous studies have validated the functional and clinical roles of the DMR associated genes in RA disease etiology, Fig. 7 and Supplemental Table S5. The RA DMRs identified in the different comparisons provides insight into potentially important epimutations (e.g., DMRs) that provide a molecular mechanism for how environmental exposures and lifestyle impact the onset of RA.

The current study identifies potential rheumatoid arthritis (RA) biomarkers with an EWAS for epimutations in control versus RA patients. An interesting observation was that the Caucasian and African American populations had distinct epimutations with increased DMR overlaps in the buccal cells. Interestingly, a combination of the Caucasian and African American samples analysis identified a DMR set that had strong overlap for both and provides a more efficient and optimal potential RA biomarker. Monocytes had DMR epimutations that were distinct from the buccal cells, but both cells had DMR associated genes previously shown to be involved in RA. Observations indicate that buccal cells can have RA DMR associated genes and provide a potential biomarker

cell for RA disease susceptibility. Larger clinical trials with a greater number of RA patients are now needed to confirm and validate the observations and DMRs identified in the current study. These expanded trials also need to identify RA susceptibility in patients without the onset of RA to explore the ability to initiate preventative therapeutic experiments to delay or prevent the onset of RA later in life. The availability of such RA susceptibility biomarkers will allow preventative medicine approaches to be considered. The current study provides the proof of concept that such RA epigenetic biomarkers may potentially be developed.

Methods summary

Clinical sample collection. Two independent single center prospective and open clinical studies were performed. The Arthritis Northwest Clinic (ANW) in Spokane, Washington, USA and Dx Biosamples, LLC in San Diego, California, USA provided samples for the current study. The participant approval and informed consent was obtained from all participants prior to the clinical sample collection. The study protocols were approved by the Quorum Review Institutional Review Board (IRB) for the ANW clinic with code # AE010831, and the Dx Biosamples company with code IORG # 0006584, and IRB # 00007904. All research was performed in accordance with relevant guidelines/regulations. The study was not designed for, nor did the IRB involve, the ability of reporting patient clinical information to be correlated. The Caucasian and African American buccal cells and Caucasian monocytes were analyzed as described in the Supplemental Methods. The buccal samples were frozen at − 20 °C, shipped on dry ice and then stored at − 80 °C prior to analysis. The monocytes were shipped immediately on ice for antibody bead isolation, as described in the Supplemental Methods, then stored at − 80 °C prior to analysis, as described in the Supplemental Methods.

Epigenetic analysis, statistics and bioinformatics. Buccal cell and monocyte DNA were isolated, as previously described[61]. Methylated DNA immunoprecipitation (MeDIP), followed by next generation sequencing (MeDIP-Seq) was performed. MeDIP-Seq, sequencing libraries, next generation sequencing, and bioinformatics analysis were performed as described[61], and are found in the Supplemental Methods. The statistical analysis and validation protocols were performed as previously described[61], and are found in the Supplemental Methods. All molecular data has been deposited into the public database at NCBI (GEO # GSE186179), and R code computational tools are available at GitHub (https://github.com/skinnerlab/MeDIP-seq) and www.skinner.wsu.edu. Lists of DMR associated genes were analyzed for functional relationships using the KEGG database (https://www.genome.jp/kegg/pathway.html) and Pathway Studio software (version 12.2.1.2: Database of functional relationships and pathways of mammalian proteins; Elsevier).

Ethics approval. The participant approval and informed consent was obtained from all participants prior to the clinical sample collection. The IRB study was approved by the Quorum Review IRB Committee for the Arthritis Northwest (ANW) Clinic, Spokane, WA, USA, with code IRB # AE010831 and Dx Biosamples, LLC, San Diego, CA, USA with code IORG # 0006584, and IRB # 00007904. All research was performed in accordance with relevant guidelines/regulations.

Data availability

All molecular data has been deposited into the public database at NCBI (GEO # GSE186179), and R code computational tools are available at GitHub (https://github.com/skinnerlab/MeDIP-seq) and www.skinner.wsu.edu.

References

1. McDougall, C., Hurd, K. & Barnabe, C. Systematic review of rheumatic disease epidemiology in the indigenous populations of Canada, the United States, Australia, and New Zealand. *Semin. Arthritis Rheum.* **46**, 675–686. https://doi.org/10.1016/j.semarthrit.2016.10.010 (2017).
2. Chaichian, Y., Genovese, M. C. & Weisman, M. H. The road to rheumatoid arthritis prevention: Challenges and opportunities. *Clin. Rheumatol.* **39**, 1379–1381. https://doi.org/10.1007/s10067-020-05016-4 (2020).
3. Lee, K. A., Min, S.-H., Kim, T.-H., Lee, S.-H. & Kim, H.-R. Magnetic resonance imaging-assessed synovial and bone changes in hand and wrist joints of rheumatoid arthritis patients. *Ann. Rheum. Dis.* **76**, 1166. https://doi.org/10.1136/annrheumdis-2017-eular.1507 (2017).
4. Daniel, C. M. *et al.* Ethnic disparities in atherosclerotic cardiovascular disease incidence and prevalence among rheumatoid arthritis patients in the united states: A systematic review. *ACR Open Rheumatol.* **2**, 525–532. https://doi.org/10.1002/acr2.11170 (2020).
5. Messina, R., Guggino, G., Benfante, A. & Scichilone, N. Interstitial lung disease in elderly rheumatoid arthritis patients. *Drugs Aging* **37**, 11–18. https://doi.org/10.1007/s40266-019-00727-z (2020).
6. Piovezana Bossolani, G. D. *et al.* Rheumatoid arthritis induces enteric neurodegeneration and jejunal inflammation, and quercetin promotes neuroprotective and anti-inflammatory actions. *Life Sci.* **238**, 116956. https://doi.org/10.1016/j.lfs.2019.116956 (2019).
7. Salliot, C., Nguyen, Y., Boutron-Ruault, M. C. & Seror, R. Environment and lifestyle: Their influence on the risk of RA. *J. Clin. Med.* https://doi.org/10.3390/jcm9103109 (2020).
8. Seror, R. *et al.* Passive smoking in childhood increases the risk of developing rheumatoid arthritis. *Rheumatology (Oxford)* **58**, 1154–1162. https://doi.org/10.1093/rheumatology/key219 (2019).
9. Bergstrom, U., Jacobsson, L. T., Nilsson, J. A., Wirfalt, E. & Turesson, C. Smoking, low formal level of education, alcohol consumption, and the risk of rheumatoid arthritis. *Scand. J. Rheumatol.* **42**, 123–130. https://doi.org/10.3109/03009742.2012.723744 (2013).
10. Gioia, C., Lucchino, B., Tarsitano, M. G., Iannuccelli, C. & Di Franco, M. Dietary habits and nutrition in rheumatoid arthritis: Can diet influence disease development and clinical manifestations?. *Nutrients* https://doi.org/10.3390/nu12051456 (2020).
11. Ilar, A., Alfredsson, L., Wiebert, P., Klareskog, L. & Bengtsson, C. Occupation and risk of developing rheumatoid arthritis: Results from a population-based case-control study. *Arthritis Care Res. (Hoboken)* **70**, 499–509. https://doi.org/10.1002/acr.23321 (2018).
12. Hart, J. E. *et al.* Ambient air pollution exposures and risk of rheumatoid arthritis. *Arthritis Care Res. (Hoboken)* **65**, 1190–1196. https://doi.org/10.1002/acr.21975 (2013).

13. Lu, B., Solomon, D. H., Costenbader, K. H. & Karlson, E. W. Alcohol consumption and risk of incident rheumatoid arthritis in women: A prospective study. *Arthritis Rheumatol.* **66**, 1998–2005. https://doi.org/10.1002/art.38634 (2014).
14. Seoane-Mato, D. *et al.* Prevalence of rheumatic diseases in adult population in Spain (EPISER 2016 study): Aims and methodology. *Reumatol. Clin.* **15**, 90–96. https://doi.org/10.1016/j.reuma.2017.06.009 (2019).
15. Sacilotto, N. C. *et al.* Real - rheumatoid arthritis in real life: Study cohort: A sociodemographic profile of rheumatoid arthritis in Brazil. *Adv. Rheumatol.* **60**, 20. https://doi.org/10.1186/s42358-020-0121-5 (2020).
16. Mochizuki, T., Ikari, K., Yano, K. & Okazaki, K. Five-year incidence of common comorbidities, such as hypertension, dyslipidemia, diabetes mellitus, cardiovascular disease, cerebrovascular disease and cancer, in older Japanese patients with rheumatoid arthritis. *Geriatr. Gerontol. Int.* **19**, 577–581. https://doi.org/10.1111/ggi.13664 (2019).
17. Plenge, R. M. Rheumatoid arthritis genetics: 2009 update. *Curr. Rheumatol. Rep.* **11**, 351–356. https://doi.org/10.1007/s11926-009-0050-0 (2009).
18. Mikhaylenko, D. S. *et al.* Genetic polymorphisms associated with rheumatoid arthritis development and antirheumatic therapy response. *Int. J. Mol. Sci.* **21**, 4911. https://doi.org/10.3390/ijms21144911 (2020).
19. Deane, K. D. *et al.* Genetic and environmental risk factors for rheumatoid arthritis. *Best Pract. Res. Clin. Rheumatol.* **31**, 3–18. https://doi.org/10.1016/j.berh.2017.08.003 (2017).
20. Okada, Y. *et al.* Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* **506**, 376–381. https://doi.org/10.1038/nature12873 (2014).
21. Guo, S. *et al.* Epigenetic regulation mediated by methylation in the pathogenesis and precision medicine of rheumatoid arthritis. *Front. Genet.* **11**, 811. https://doi.org/10.3389/fgene.2020.00811 (2020).
22. Nilsson, E., Sadler-Riggleman, I. & Skinner, M. K. Environmentally induced epigenetic transgenerational inheritance of disease. *Environ. Epigenetics* **4**, 1–13. https://doi.org/10.1093/eep/dvy016 (2018).
23. Hammaker, D. & Firestein, G. S. Epigenetics of inflammatory arthritis. *Curr. Opin. Rheumatol.* **30**, 188–196. https://doi.org/10.1097/BOR.0000000000000471 (2018).
24. Frank-Bertoncelj, M., Klein, K. & Gay, S. Interplay between genetic and epigenetic mechanisms in rheumatoid arthritis. *Epigenomics* **9**, 493–504. https://doi.org/10.2217/epi-2016-0142 (2017).
25. Vecellio, M., Wu, H., Lu, Q. & Selmi, C. The multifaceted functional role of DNA methylation in immune-mediated rheumatic diseases. *Clin. Rheumatol.* https://doi.org/10.1007/s10067-020-05255-5 (2020).
26. Calle-Fabregat, C., Morante-Palacios, O. & Ballestar, E. Understanding the relevance of DNA methylation changes in immune differentiation and disease. *Genes* https://doi.org/10.3390/genes11010110 (2020).
27. Chen, S. *et al.* Genome-wide DNA methylation profiles reveal common epigenetic patterns of interferon-related genes in multiple autoimmune diseases. *Front. Genet.* **10**, 223. https://doi.org/10.3389/fgene.2019.00223 (2019).
28. Anaparti, V., Agarwal, P., Smolik, I., Mookherjee, N. & El-Gabalawy, H. Whole blood targeted bisulfite sequencing and differential methylation in the C6ORF10 gene of patients with rheumatoid arthritis. *J. Rheumatol.* **47**, 1614–1623. https://doi.org/10.3899/jrheum.190376 (2020).
29. Zouali, M. DNA methylation signatures of autoimmune diseases in human B lymphocytes. *Clin. Immunol.* **222**, 108622. https://doi.org/10.1016/j.clim.2020.108622 (2020).
30. Liebold, I. *et al.* Peripheral blood mononuclear cells are hypomethylated in active rheumatoid arthritis and methylation correlates with disease activity. *Rheumatology (Oxford)* https://doi.org/10.1093/rheumatology/keaa649 (2020).
31. Nygaard, G. & Firestein, G. S. Restoring synovial homeostasis in rheumatoid arthritis by targeting fibroblast-like synoviocytes. *Nat. Rev. Rheumatol.* **16**, 316–333. https://doi.org/10.1038/s41584-020-0413-5 (2020).
32. Karami, J. *et al.* Epigenetics in rheumatoid arthritis; fibroblast-like synoviocytes as an emerging paradigm in the pathogenesis of the disease. *Immunol. Cell Biol.* **98**, 171–186. https://doi.org/10.1111/imcb.12311 (2020).
33. Karouzakis, E., Gay, R. E., Gay, S. & Neidhart, M. Epigenetic control in rheumatoid arthritis synovial fibroblasts. *Nat. Rev. Rheumatol.* **5**, 266–272. https://doi.org/10.1038/nrrheum.2009.55 (2009).
34. Karouzakis, E. *et al.* Analysis of early changes in DNA methylation in synovial fibroblasts of RA patients before diagnosis. *Sci. Rep.* **8**, 7370. https://doi.org/10.1038/s41598-018-24240-2 (2018).
35. Ospelt, C., Gay, S. & Klein, K. Epigenetics in the pathogenesis of RA. *Semin. Immunopathol.* **39**, 409–419. https://doi.org/10.1007/s00281-017-0621-5 (2017).
36. Grennan, D. M., Sanders, P. A., Thomson, W. & Dyer, P. A. Rheumatoid arthritis: Inheritance and association with other autoimmune diseases. *Dis. Mark.* **4**, 157–162 (1986).
37. Hajalilou, M. *et al.* Familial rheumatoid arthritis in patients referred to rheumatology clinics of Tabriz, Iran. *Int. J. Rheum. Dis.* **15**, 110–115. https://doi.org/10.1111/j.1756-185X.2011.01664.x (2012).
38. Blunk, I. *et al.* Genomic imprinting analyses identify maternal effects as a cause of phenotypic variability in type 1 diabetes and rheumatoid arthritis. *Sci. Rep.* **10**, 11562. https://doi.org/10.1038/s41598-020-68212-x (2020).
39. Anway, M. D., Cupp, A. S., Uzumcu, M. & Skinner, M. K. Epigenetic transgenerational actions of endocrine disruptors and male fertility. *Science* **308**, 1466–1469. https://doi.org/10.1126/science.1108190 (2005).
40. Wu, H. *et al.* Epigenetics as biomarkers in autoimmune diseases. *Clin. Immunol.* **196**, 34–39. https://doi.org/10.1016/j.clim.2018.03.011 (2018).
41. Ciechomska, M., Roszkowski, L. & Maslinski, W. DNA methylation as a future therapeutic and diagnostic target in rheumatoid arthritis. *Cells* **8**, 953. https://doi.org/10.3390/cells8090953 (2019).
42. Rodriguez-Ubreva, J. *et al.* Inflammatory cytokines shape a changing DNA methylome in monocytes mirroring disease activity in rheumatoid arthritis. *Ann. Rheum. Dis.* **78**, 1505–1516. https://doi.org/10.1136/annrheumdis-2019-215355 (2019).
43. Deane, K. D. & Holers, V. M. Rheumatoid arthritis: Pathogenesis, prediction and prevention: An emerging paradigm shift. *Arthritis Rheumatol.* https://doi.org/10.1002/art.41417 (2020).
44. Mahler, M., Martinez-Prat, L., Sparks, J. A. & Deane, K. D. Precision medicine in the care of rheumatoid arthritis: Focus on prediction and prevention of future clinically-apparent disease. *Autoimmun. Rev.* **19**, 102506. https://doi.org/10.1016/j.autrev.2020.102506 (2020).
45. Dagur, P. K. & McCoy, J. P. Jr. Collection, storage, and preparation of human blood cells. *Curr. Protoc. Cytom.* **73**, 1–16. https://doi.org/10.1002/0471142956.cy0501s73 (2015).
46. Garrido, N. *et al.* Sperm DNA methylation epimutation biomarker for paternal offspring autism susceptibility. *Clin. Epigenetics* **13**, 6. https://doi.org/10.1186/s13148-020-00995-2 (2021).
47. Ferreira, H. B., Melo, T., Paiva, A. & Domingues, M. D. R. Insights in the role of lipids, oxidative stress and inflammation in rheumatoid arthritis unveiled by new trends in lipidomic investigations. *Antioxidants.* https://doi.org/10.3390/antiox10010045 (2021).
48. Garnero, P., Landewe, R. & Chapurlat, R. D. The role of biochemical markers of joint tissue remodelling to predict progression and treatment efficacy in inflammatory rheumatic diseases. *Rheumatology (Oxford)* **59**, 1207–1217. https://doi.org/10.1093/rheumatology/kez647 (2020).
49. Drobinski, P. J., Bay-Jensen, A. C., Karsdal, M. A., Sardar, S. & Siebuhr, A. S. Connective tissue remodelling is differently modulated by tocilizumab versus methotrexate monotherapy in patients with early rheumatoid arthritis: The AMBITION study. *Arthritis Res. Ther.* **23**, 13. https://doi.org/10.1186/s13075-020-02378-7 (2021).

50. Sammons, S., Sedrak, M. S. & Kimmick, G. G. The evolving complexity of treating hormone receptor-positive, human epidermal growth factor receptor-2 (HER2)-negative breast cancer: Special considerations in older breast cancer patients-part i: Early-stage disease. *Drugs Aging* **37**, 331–348. https://doi.org/10.1007/s40266-020-00748-z (2020).

51. Erickson, R. P. & Mitchison, N. A. The low frequency of recessive disease: Insights from ENU mutagenesis, severity of disease phenotype, GWAS associations, and demography: An analytical review. *J. Appl. Genet.* **55**, 319–327. https://doi.org/10.1007/s13353-014-0203-3 (2014).

52. Yamashita, S. *et al.* Genetic and epigenetic alterations in normal tissues have differential impacts on cancer risk among tissues. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 1328–1333. https://doi.org/10.1073/pnas.1717340115 (2018).

53. Pigeyre, M., Yazdi, F. T., Kaur, Y. & Meyre, D. Recent progress in genetics, epigenetics and metagenomics unveils the pathophysiology of human obesity. *Clin. Sci. (Lond)* **130**, 943–986. https://doi.org/10.1042/CS20160136 (2016).

54. Gonzalez-Jaramillo, V. *et al.* The role of DNA methylation and histone modifications in blood pressure: A systematic review. *J. Hum. Hypertens.* **33**, 703–715. https://doi.org/10.1038/s41371-019-0218-7 (2019).

55. Collier, A. Y. *et al.* Racial and ethnic representation in epigenomic studies of preterm birth: A systematic review. *Epigenomics* https://doi.org/10.2217/epi-2020-0007 (2020).

56. Culemann, S., Gruneboom, A. & Kronke, G. Origin and function of synovial macrophage subsets during inflammatory joint disease. *Adv. Immunol.* **143**, 75–98. https://doi.org/10.1016/bs.ai.2019.08.006 (2019).

57. Salliot, C., Bombardier, C., Saraux, A., Combe, B. & Dougados, M. Hormonal replacement therapy may reduce the risk for RA in women with early arthritis who carry HLA-DRB1 *01 and/or *04 alleles by protecting against the production of anti-CCP: Results from the ESPOIR cohort. *Ann. Rheum. Dis.* **69**, 1683–1686. https://doi.org/10.1136/ard.2009.111179 (2010).

58. Hussin, D., Shaat, R. M., Metwally, S. S. & Awad, M. The significance of serum 14–3-3eta level in rheumatoid arthritis patients. *Clin. Rheumatol.* https://doi.org/10.1007/s10067-020-05524-3 (2021).

59. Gibson, D. S. *et al.* Anti-tumour necrosis factor-alpha response associated with combined CD226 and HLA-DRB1*0404 haplotype in rheumatoid arthritis. *Clin Exp Rheumatol* (2021).

60. Wu, Y. *et al.* Serum 14–3-3eta is a marker that complements current biomarkers for the diagnosis of RA: Evidence from a meta-analysis. *Immunol. Invest.* https://doi.org/10.1080/08820139.2020.1817069 (2020).

61. King, S. E. *et al.* Sperm epimutation biomarkers of obesity and pathologies following DDT induced epigenetic transgenerational inheritance of disease. *Environ. Epigenet* **5**, 1–15. https://doi.org/10.1093/eep/dvz008 (2019).

Acknowledgements

Author contributions

G.C. patient's recruitment, clinical and sample collection oversight, data analysis, editing manuscript. H.K. patient's recruitment, clinical and sample collection oversight, data analysis, editing manuscript. I.S.R. molecular analysis, data analysis, editing manuscript. D.B. bioinformatics, data analysis, editing manuscript. E.N. sample processing, data analysis, editing manuscript. M.K.S. conceived, data analysis, funding acquisition, wrote and edited manuscript.

Funding

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-021-03170-6.

Correspondence and requests for materials should be addressed to M.K.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.