

ARTICLE OPEN



Machine learning-based detection of immune-mediated diseases from genome-wide cell-free DNA sequencing datasets

Huiwen Che¹, Tatjana Jatsenko¹, Lore Lannoo², Kate Stanley¹, Luc Dehaspe³, Leen Vancoillie³, Nathalie Brison³, Ilse Parijs³, Kris Van Den Bogaert³, Koenraad Devriendt³, Sabien Severi⁴, Ellen De Langhe^{4,5}, Severine Vermeire^{6,7}, Bram Verstockt^{6,7}, Kristel Van Calsteren² and Joris Robert Vermeesch^{1,3}✉

The early detection of tissue and organ damage associated with autoimmune diseases (AID) has been identified as key to improve long-term survival, but non-invasive biomarkers are lacking. Elevated cell-free DNA (cfDNA) levels have been observed in AID and inflammatory bowel disease (IBD), prompting interest to use cfDNA as a potential non-invasive diagnostic and prognostic biomarker. Despite these known disease-related changes in concentration, it remains impossible to identify AID and IBD patients through cfDNA analysis alone. By using unsupervised clustering on large sets of shallow whole-genome sequencing (sWGS) cfDNA data, we uncover AID- and IBD-specific genome-wide patterns in plasma cfDNA in both the obstetric and general AID and IBD populations. We demonstrate that pregnant women with AID and IBD have higher odds of receiving inconclusive non-invasive prenatal screening (NIPS) results. Supervised learning of the genome-wide patterns allows AID prediction with 50% sensitivity at 95% specificity. Importantly, the method has the potential to identify pregnant women with AID during routine NIPS. Since AID pregnancies have an increased risk of severe complications, early recognition or detection of new-onset AID can redirect pregnancy management and limit potential adverse events. This method opens up new avenues for screening, diagnosis and monitoring of AID and IBD.

npj Genomic Medicine (2022)7:55; <https://doi.org/10.1038/s41525-022-00325-w>

INTRODUCTION

Increased levels of cell-free DNA (cfDNA) have been observed in individuals with systemic lupus erythematosus (SLE), a common and prototypical autoimmune disease (AID), as early as the 1960s¹. The elevation of cfDNA has been found in other AID, such as rheumatoid arthritis (RA)² and inflammatory bowel disease (IBD), including Crohn's disease (CD) and ulcerative colitis (UC)^{3,4}. Despite the existence of different predictive and prognostic biomarkers for AID and IBD, more robust non-invasive markers to deliver precision medicine are needed^{5,6}. Given the observed elevation, cfDNA has become a promising biomarker for prediction, monitoring, and treatment response stratification for AID and IBD^{7–11}. Although assessment of overall cfDNA concentrations in these diseases allows for the inference of AID and IBD activity, it lacks specificity and provides no information about qualitative disease-associated cfDNA changes, which might aid in disease diagnosis or prognosis.

CfDNA analysis has been clinically implemented for prenatal¹² and cancer¹³ management. Non-invasive prenatal screening (NIPS) for aneuploidies detection has become a cornerstone of prenatal care over the past decade¹⁴. Placenta-derived cfDNA is present in maternal plasma and serum¹⁵, and different approaches have been developed to detect fetal chromosomal anomalies^{12,16}. Despite the high accuracy of NIPS, approximately 0.2–5.4% of NIPS remain inconclusive, even after repeated testing^{17–21}. Multiple studies have reported that the most common reason for inconclusive NIPS results is a low fetal fraction (FF), accounting for up to 50% of test failures^{22–24}. A low FF can be due to either a

low amount of placental DNA (e.g. early gestational age) or to an increased amount of maternal DNA. A number of maternal conditions can increase the maternal cfDNA fraction^{22,25}. Recently, AID has been identified as a contributor to reduced FF^{22,26–28}. As a consequence, pregnant AID patients have significantly higher odds of receiving an inconclusive NIPS result when compared to the general obstetric population^{24,26}.

The growing volume of shallow whole-genome cfDNA sequencing (sWGS) data permits the exploration of disease-associated cfDNA signatures using machine learning approaches. By applying GIPX²⁹, our recently developed algorithm for sWGS data mining using unsupervised clustering and supervised machine learning, we have already demonstrated the presence of cancer-specific patterns in plasma cfDNA profiles from cancer patients. Therefore, we hypothesized that autoimmune-related disease patterns might equally be elucidated via this approach. By interrogating over 80,000 cfDNA profiles from pregnant women and an independent cohort of non-pregnant individuals, we revealed AID- and IBD-associated genome-wide cfDNA signatures that allow the detection of the disease both during pregnancy and in non-pregnant patients. This study demonstrates that sWGS of cfDNA can identify women with immune-mediated diseases early during pregnancy, having a potential utility to detect, diagnose and monitor individuals at high risk. Our approach can be easily integrated with routine NIPS workflows and pending success in large-scale prospective trials can be broadly implemented for better prenatal management.

¹Department of Human Genetics, Laboratory for Cytogenetics and Genome Research, KU Leuven, Leuven, Belgium. ²Department of Gynecology and Obstetrics, University Hospitals Leuven, Leuven, Belgium. ³Centre for Human Genetics, University Hospitals Leuven, Leuven, Belgium. ⁴Department of Rheumatology, University Hospitals Leuven, Leuven, Belgium. ⁵Department of Development and Regeneration, Laboratory of Tissue Homeostasis and Disease, KU Leuven, Leuven, Belgium. ⁶Department of Gastroenterology and Hepatology, University Hospitals Leuven, Leuven, Belgium. ⁷Department of Chronic Diseases and Metabolism, TARGID-IBD, KU Leuven, Leuven, Belgium. ✉email: joris.vermeesch@uzleuven.be

RESULTS

AID and IBD are overrepresented in inconclusive NIPS and are characterized by recurrent cfDNA profile patterns

We analyzed 81,611 cfDNA samples from pregnant women and identified a total 185 inconclusive cases (0.23%) using an unbiased NIPS analysis pipeline coined GIPSeq¹⁶. The inconclusive result (Cohort I) was attributed to either a high variance in genome-wide cfDNA read counts (high quality score (QS); $n = 143$; 77%) or a low fetal fraction (low FF; $n = 42$; 23%). Since all the inconclusive cases had a second sampling, and for some, an additional third technical replicate, a total of 406 cfDNA profiles were obtained from the 185 pregnant women.

Using GIPXplore, we performed explorative analysis on the set of inconclusive NIPS profiles and randomly selected conclusive NIPS control samples. Two to three conclusive control samples were matched to the sample processing time, sequencing batch and library preparation kit of inconclusive cases, accumulating to 1024 samples. The unsupervised clustering revealed a separation between a subset of the inconclusive profiles with deviating QS and conclusive NIPS profiles (Fig. 1a). Most of the inconclusive profiles with deviating QS fell into two large groups, where clusters 4, 15, and 18 formed one group and clusters 7, 8 and 17 composed another one with similar genome-wide profiles (Supplementary Fig. 1). A few scattered cases could be found in clusters 13, 19 and 20. The inconclusive profiles with low FF, in general, overlapped with conclusive profiles. Repeated samples, both technical and biological replicates, from the same inconclusive case, showed highly concordant profiles, with 74% of repeats located in the same cluster or having close (<5% of the embedding dimensions coordinates limits) geometric coordinates in the tSNE representation. There were 48 (26%) cases that exhibited less similar profiles among the independent NIPS repeats, which may indicate dynamic changes in cfDNA or transient status during pregnancies. To eliminate potential biased observation due to the inclusion of biological and technical replicates, analysis repeated with only the first sample from the series revealed a similar clustering pattern (Supplementary Fig. 2).

We hypothesized that the observed clusters might reflect different maternal and/or fetal pathologies or conditions. To explore this, we retrospectively reviewed the maternal clinical records for diagnoses of immune-mediated diseases. As presented in Fig. 1b, 17% ($n = 25$) of inconclusive cases owing to a high QS were found to have AID, including SLE, antiphospholipid syndrome (APS), hypothyroidism, primary Sjogren's syndrome, Graves' disease, RA, psoriasis, and spondyloarthritis. Among these 25, 13 women had SLE, of which 3 had APS, 1 had thrombosis, and 1 had psoriasis as a comorbidity. The estimated prevalence of SLE is 134.5 cases per 100,000 persons in the Caucasian population^{30,31}. Based on this prevalence rate and the number of identified SLE cases, pregnant women with SLE had at least a sixty-fold higher chance of an inconclusive NIPS result than those without SLE (unadjusted odds ratio [OR] 63.4). In addition, 4% of cases ($n = 5$) had IBD. Given that the prevalence is 312 per 100,000 in women giving birth³², pregnant women with IBD have at least a 9-fold higher chance of an inconclusive NIPS result (unadjusted OR = 9.0) based on our cohort. Finally, 5% cases ($n = 7$) had blood disorders. In the low FF group, five (12%) inconclusive cases were associated with AID and two (5%) with blood disorders, respectively. Cluster annotation revealed a dominant representation of AID cases in cluster 4 (Fig. 1c). Investigation of only the first sample from inconclusive cases also demonstrated a high genome-wide similarity among a subset of AID ($n = 10$), particularly SLE (Supplementary Fig. 3). With that, 5 cases with unknown or healthy medical history, one case with IBD and another case with a blood disorder shared a common lineage with these AID cases (for simplicity, referred to as the AID clusters). Taken together, GIPXplore analysis of the inconclusive cases identified a

cluster of pregnancies enriched with AID ($P < 1e-06$). In the remaining clusters (13, 19 and 20), no common cause could be identified in the clinical records.

Next, we questioned whether the AID/IBD-specific signal could also be observed in pregnancies with a conclusive NIPS result. To explore this, we combined the cfDNA profiles of 44 inconclusive NIPS cases (96 samples) that were included in the aforementioned analysis from the Cohort I and additionally Cohort II - 69 conclusive NIPS cases from pregnant women with known diagnoses of AID (SLE, APS, hypothyroidism/Hashimoto's disease, hyperthyroidism/Graves' disease) or IBD (CD and UC; Fig. 2a). GIPXplore on these cases together with control NIPS samples showed that disease-associated inconclusive NIPS profiles drifted away from the controls and formed sub-clusters as expected (Fig. 2b and Supplementary Fig. 4). More importantly, 4 SLE and 1 IBD cases that had a conclusive NIPS result co-localized with inconclusive cases in clusters 10, 13 and 7, respectively (Fig. 2c).

GIPXplore enables AID prediction using NIPS profiles

Given that SLE is a prototypical autoimmune disease that is highly relevant to pregnancy management, we examined whether the genome-wide patterns observed in SLE could be used to identify pregnancies affected with SLE and/or pregnancies affected with other AID. We established a classifier trained on SLE cases and tested it using obstetric cfDNA data with and without disease conditions. We used a support vector machine (SVM) classifier based on genome-wide NIPS features. For this analysis, we grouped together SLE cases with both inconclusive (Cohort I) and conclusive (Cohort II) NIPS. SLE cases from Cohort I had more than one biological replicate. We split the NIPS cohort into the training set that included 12 inconclusive (Cohort I) and 22 conclusive (Cohort II) SLE cases, and the validation set ($n = 399$) that included the second biological repeat sample (Cohort I, $n = 13$) from each case in the training set (Fig. 3). First, we performed a predictive analysis in the training set to distinguish SLE from control NIPS profiles. Using leave-one-out cross-validation with pre-selected parameters, the estimated sensitivity was 71% at a specificity of 95%. The receiver operating characteristic (ROC) analysis showed an area under the curve (AUC) value of 0.91. A final classifier was then built on the training set. Applying the classifier on the independent validation set, all repeat SLE samples were correctly predicted. We also applied the final model to an independent set of cases with other AID, IBD and blood disorders. In concordance with the clustering analysis, where these cases could group with the SLE cases, 33%, 38% and 44% of profiles that primarily presented with other AID, IBD and blood disorders were predicted as SLE-like profiles. The specificities estimated in the training and validation set were consistent.

Shared cfDNA profiles signatures in pregnant and non-pregnant populations with AID and IBD

We hypothesized that the disease-associated cfDNA genome-wide patterns observed in the obstetric population may also be characteristic to the general AID and IBD patient population. To test this, we investigated 24 SLE and 33 CD cfDNA samples from non-pregnant patients (Cohort III; Supplementary Data). The 24 non-pregnant SLE patients include 21 females and 3 males. Among 33 non-pregnant CD patients, 26 are males. We first performed the clustering analysis on these samples together with 104 cfDNA profiles from non-pregnant controls, in which two distinct disease-like clusters (4 and 7) formed, suggesting not only the presence of disease-associated cfDNA genome-wide patterns, but also potential stratification of SLE and CD profiles (Supplementary Fig. 5a, b). The separation of the clusters was not driven by sex (Supplementary Fig. 5c).

Next, using unsupervised clustering analysis, we assessed the similarity of disease-associated profiles in both non-pregnant and

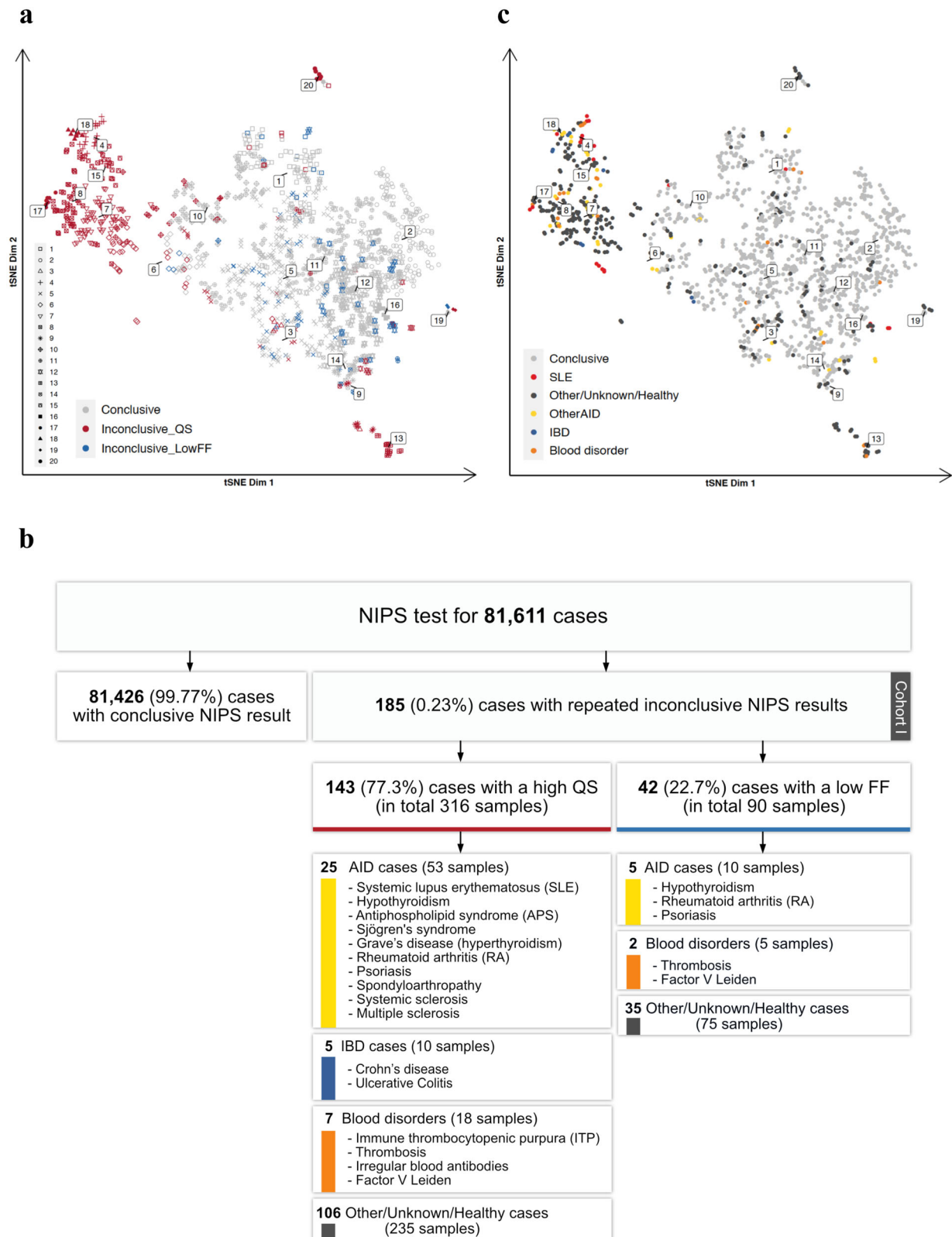
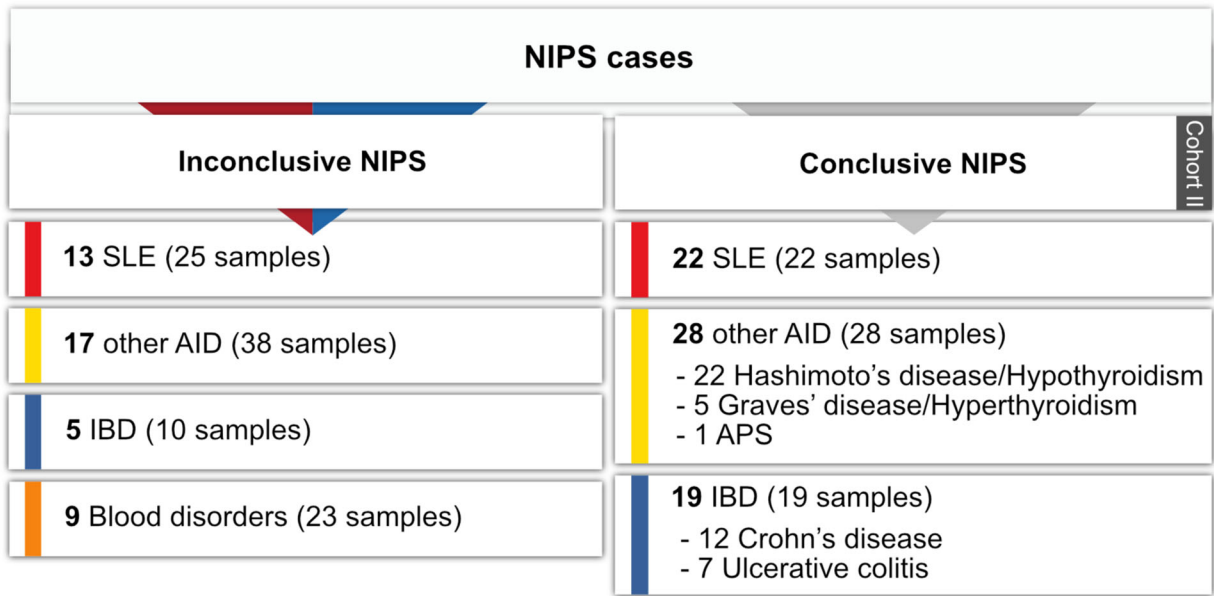


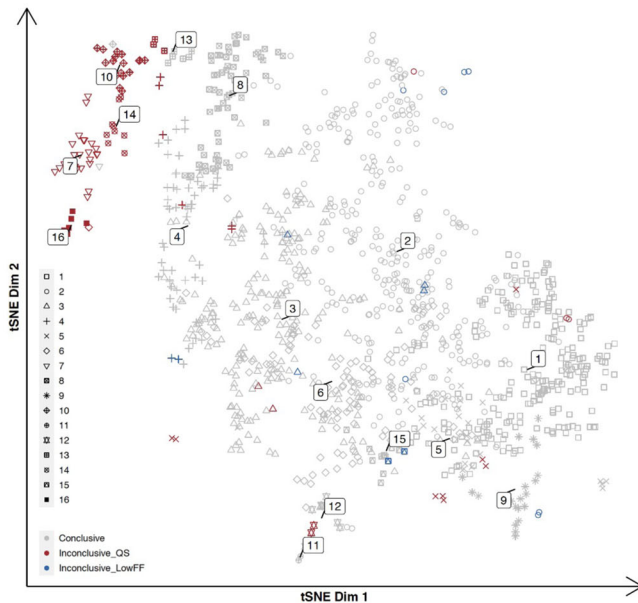
Fig. 1 Clustering of repeated inconclusive NIPS. **a** tSNE representation of inconclusive samples ($n = 406$) and conclusive NIPS controls ($n = 1024$). Each point represents one sample. Colours red and blue indicate inconclusive samples due to deviating QS and low FF respectively. Point shape represents the clusters being identified using Walktrap community detection. **b** overview of the study cohort and cases with immune-mediated diseases. **c** tSNE representation with clinical information annotation.

a



Cohort II

b



c

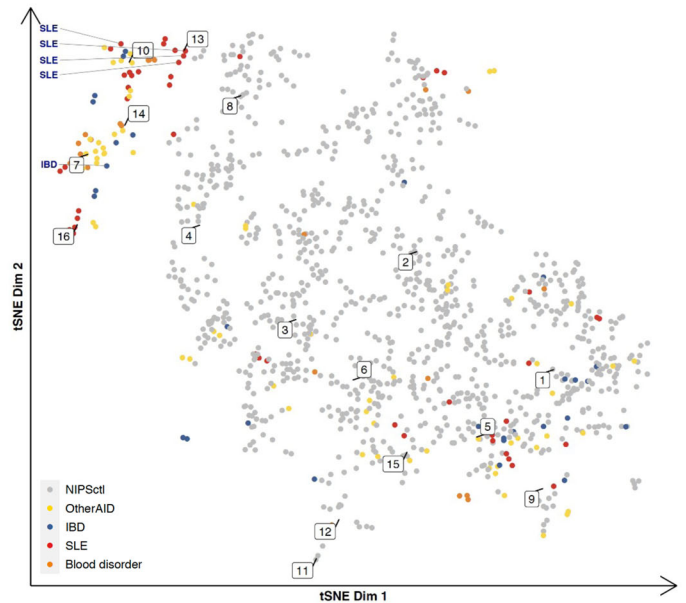


Fig. 2 Analysis of AID and IBD from both inconclusive and conclusive NIPS. a summary of analysis cohort. **b** tSNE representation of the cohort, including AID and IBD NIPS with inconclusive and conclusive results. Point color indicates NIPS result. **c** the same tSNE representation as in **b**, with the annotation of phenotypes shown in different colors. NIPSetI represents NIPS samples with conclusive results. Five conclusive SLE and IBD NIPS cases that are colocalized with inconclusive AID and IBD samples are annotated additionally with text.

pregnant subjects. Interestingly, the unsupervised clustering revealed co-clustering of pregnant and non-pregnant subjects with AID and IBD (Fig. 4a, b). Seven non-pregnant SLE and eight non-pregnant CD cases were found in disease-associated clusters. Analogous to the SLE-associated clusters (Supplementary Fig. 4, clusters 13 and 16) identified in the previous analysis, both pregnant and non-pregnant SLE profiles assembled in clusters 11 and 17 (Fig. 4b). The shared cfDNA signatures in pregnant and non-pregnant patients presenting with AID and IBD were also

evident in clusters 15, 18, 21 and 22. Moreover, one of the non-pregnant CD cases, clustered with AID in cluster 18, having AID (spondyloarthritis) comorbidity. Two other non-pregnant CD cases in cluster 3, grouping with AID, have other AID comorbidities (one with psoriasis and the other with autoimmune hemolytic anemia) as well. Other non-pregnant CD cases that clustered with AID were not associated with AID comorbidities. Cluster 21 was notably divergent from other AID and IBD clusters in multiple replicates of tSNE representations. Similar results were obtained in the cluster

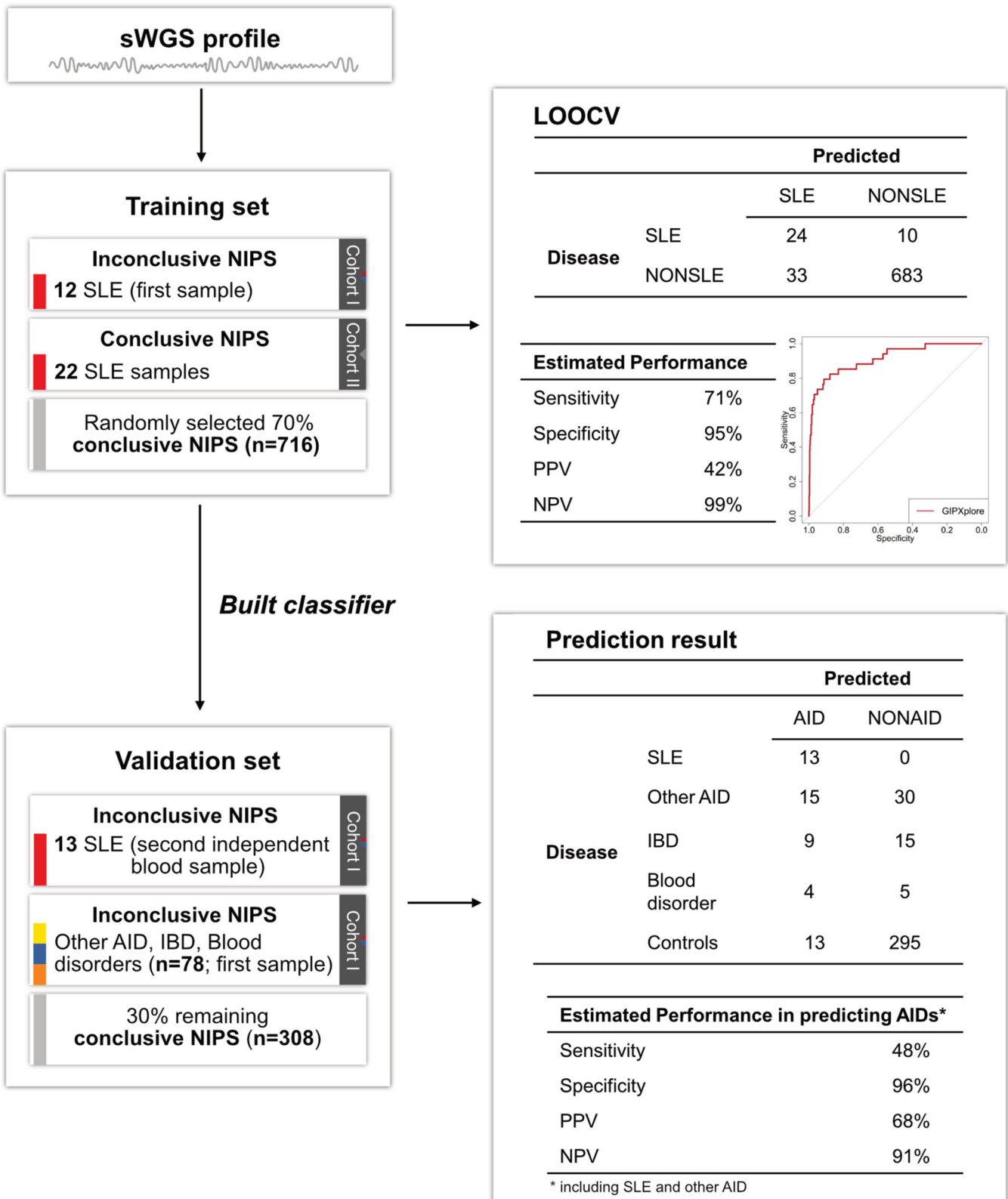


Fig. 3 Workflow for building the classifier for AID prediction and the resulting performance. As one of the inconclusive SLE cases is the second pregnancy of the existing inconclusive case, we only included 12 inconclusive SLE cases in the training set. The one left out was used in the validation set as a 'second' SLE sample. In the prediction result for the validation set, performance characteristics (sensitivity, specificity, positive predictive value - PPV, negative predictive value - NPV) were calculated only using prediction results from AID (SLE and other AID) and controls.

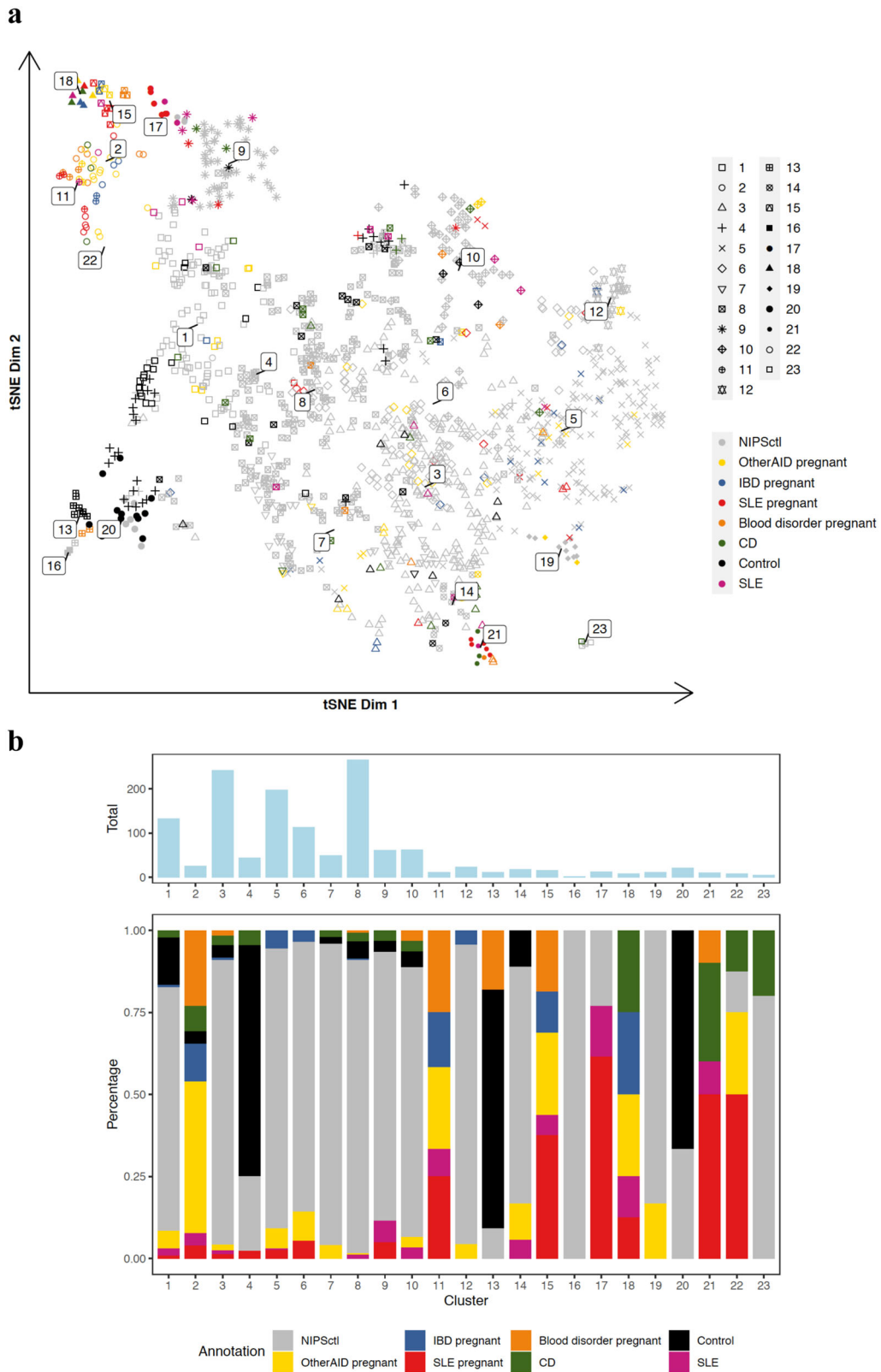


Fig. 4 Clustering of cfDNA profiles from pregnant and non-pregnant AID and IBD. a tSNE representation of the analysis cohort. Point shape indicates the clusters that were defined by the Walktrap algorithm. Point color indicates phenotypic information. Control represents non-pregnant control samples. **b** upper bar plot shows the number of samples in each cluster and lower bar plot shows composition of samples in each cluster.

analysis excluding non-pregnant control subjects (Supplementary Fig. 6). We did not uncover any clinical factors that could have led to the separation using available information.

Though the classifier for AID prediction was not built for the general population, we reasoned that the disease-associated changes in the genome-wide profile of non-pregnant subjects could also be identified by the built classifier. When we applied the trained model on this non-pregnant cohort, a positive likelihood ratio (LR+) of 13 and negative likelihood ratio (LR-) of 0.52 for SLE, and a LR+ of 7.09 and LR- of 0.76 for CD were obtained (Supplementary Table 1). In the subsequent linear regression analysis, for non-pregnant SLE patients with high SLEDAI (Systemic Lupus Erythematosus Disease Activity Index), the odds of correct prediction increased by 17-fold (adjusted OR = 17.7), after adjusting for patient age, disease activity, and the existence of comorbidity. Disease activity is not significantly associated with the prediction outcome, after adjusting for the patient's age, SLEDAI and the existence of comorbidity.

DISCUSSION

We demonstrate that the genome-wide cfDNA patterns of a subset of AID- and IBD-affected individuals can be distinguished from controls, and such cfDNA patterns can be used for the classification of immune-mediated diseases. The early detection of damage caused by SLE and other AID has been identified as key to improve long-term survival³³. The potential to use cfDNA patterns as a non-invasive biomarker for immune-mediated diseases identification, paves the way towards screening, diagnosis and monitoring of AID and IBD. Since the patterns are similar in both pregnant and non-pregnant individuals and the disease-specific patterns can be identified during the first trimester of pregnancy, this opens the potential to predict the presence of AID and IBD following NIPS.

The observed similarity in cfDNA patterns across clinically distinct AID and IBD may reflect similarities in the inflammatory processes and/or cfDNA release and clearance mechanisms involved in these diseases. AID cause inflammation and damage to various tissues, often characterized by increased apoptosis and impaired clearance of apoptotic cells¹¹. The impaired clearance of the cell death process can affect cfDNA levels and engage intracellular antigens that lead to auto-inflammatory responses³⁴. The potential role of cfDNA in the pathogenesis of IBD has been previously suggested¹⁰. Multiple lines of evidence have suggested that cfDNA is mainly released via apoptosis in healthy individuals^{35,36}. However, different types of programmed cell death, such as NETosis and pyroptosis, can also contribute to elevated cfDNA levels^{9,11,37,38}. Particularly, NETosis, where neutrophils extrude DNA into the extracellular environment in the form of neutrophil extracellular traps (NETs)³⁹, was found to be associated with the elevation of cfDNA levels in pathological conditions such as SLE and RA^{40,41}. It has also been observed that NETs play an important role in IBD^{42,43}, as well as other autoimmune and inflammatory comorbidities, including psoriasis, primary Sjogren's syndrome and Hashimoto's thyroiditis^{11,44–46}. Interestingly, neutrophil NETs were found to be associated with coagulation and thus promote thrombosis⁴⁷. In line with this, we demonstrated that cfDNA profiles with thrombosis appeared to colocalize with autoimmune diseases. Besides the release of cfDNA, the abnormal activity of DNase I and DNase I-like III (DNase IL3), two nucleases that carry out the digestion of nucleosomal DNA in plasma, have been associated with SLE and IBD^{48–51}. Chan et al. (2014) have demonstrated that plasma cfDNA from SLE patients with an active disease has a distinct cfDNA profile characterized by fragment size shortening, as well as aberrant genomic and methylation patterns⁴⁸. The level of abnormalities was correlated with anti-double-stranded DNA (anti-dsDNA) antibody levels, suggesting a potential association with the impaired clearance

of cfDNA. However, the role of cfDNA in these complex diseases is yet to be fully explored, and the cause of shared qualitative cfDNA patterns in SLE and IBD patients observed in our study is not yet understood. Furthermore, SLE patients are more likely to have an altered genome-wide cfDNA profile with more heterogeneous subclusters than other autoimmune and inflammatory diseases. Though the numbers are small, interrogating non-pregnant SLE cases showed a putative association between cfDNA profile aberrations and a high Systemic Lupus Erythematosus Disease Activity Index (SLEDAI). Whether the aberration is the result of disease activity or pathogenic mechanisms requires further examination.

We demonstrate that AID and IBD during pregnancy are risk factors for an inconclusive NIPS. While AID and IBD may be associated with impaired cfDNA degradation process, a higher maternal cell turn-over resulting in low fetal fractions may also result in an inconclusive NIPS. Although other reports indeed showed AID to be a risk factor for inconclusive NIPS as a consequence of a reduced fetal fraction, we show that 83% of AID-related inconclusive NIPs are actually due to atypical cfDNA sequencing profiles. While no study has investigated the association between inconclusive NIPS and IBD, we show that IBD may also exhibit altered cfDNA patterns. The potential to predict the presence of AID and IBD during the first trimester of pregnancy offers new opportunities for prenatal management following NIPS. Pregnancies with AID, particularly SLE, are considered high-risk for both mother and fetus. These risks include disease flares, pre-eclampsia, miscarriages, preterm birth, and intrauterine growth restriction^{52,53}. New-onset SLE may develop during pregnancy, in which the detection can be complicated by pregnancy-related changes and the clinical manifestation might be more severe⁵⁴. Women with known AID/IBD are referred to a high-risk obstetric unit with multidisciplinary care and an intensive pregnancy follow-up.

Physiological changes, such as pregnancy, can lead to a distinct sWGS cfDNA profile (Supplementary Fig. 7). In the context of AID and IBD, the pathological signals related to the presence of the disease appear to be dominant. That is, both pregnant and non-pregnant individuals with AID and IBD cluster together and are similarly classified, suggesting that the disease-associated cfDNA signatures override the pregnancy-related signatures. This is likely due to the fact that the majority of DNA fragments in maternal plasma are maternally derived. In both pregnant and non-pregnant cases, AID and IBD may show either altered or normal-like cfDNA profiles. It is unclear at this stage whether cases with altered profiles have worse outcomes in terms of pregnancy, treatment response, disease activity or other clinical parameters. Studies have investigated the impact of maternal medications (e.g. immunomodulatory therapy) on NIPS results, but there is no clear consensus. Some studies suggested that the use of immunomodulators may reduce FF and/or increase variance and thereby have an adverse effect on NIPS performance^{55,56}. Other studies suggest that immunomodulators suppress disease activity and could therefore be expected to normalize NIPS profiles^{24,27}.

Given the design of the current study, the clinical information for the inconclusive NIPS cohort is incomplete. We only partially unraveled the phenotypes linked to the clusters being identified. Future analysis of maternal conditions, and their impact on fetal health and pregnancy outcomes require a standardized model with systematic clinical data collection. This would allow further evaluation of the effects of AID and IBD on cfDNA patterning and its potential use in prenatal care. For early diagnosis in presymptomatic AID/IBD individuals at risk (e.g. first-degree relatives in multiplex IBD families)⁵⁷, cfDNA genomic profiles might have a broader place as non-invasive biomarkers. Age- and gender-matched case and control cohort needs to be evaluated to eliminate potential bias. While we examined cfDNA samples collected from non-pregnant patients with different treatment

and status of AID and IBD disease, collecting longitudinal cfDNA samples before and during the treatment course can shed more light on the association between the treatment and change of cfDNA profile. Large, prospective cohort studies are therefore warranted, which could explore its prognostic and predictive accuracy in immune-mediated disorders aiming to further individualize treatment strategies.

In conclusion, the ability to identify AID and IBD with cfDNA sequence patterns creates new opportunities for immune-mediated disease management. More importantly, using this approach during NIPS can identify pregnant women with higher risks of pregnancy complications and redirect prenatal management.

METHODS

Study design and participants

The study retrospectively investigated NIPS data from a population of 81,611 pregnant women between July 2017 and December 2020. Peripheral blood was sampled in pregnant women from the 10th week of gestation onwards in Cell-Free DNA collection tubes (Roche Diagnostics). For the cases with an inconclusive result, a second blood sample was requested. Medical history of these repeated inconclusive cases was obtained by reviewing medical records from the referring clinical team with consent from 185 individuals. The study was approved by University Hospitals Leuven Ethics Committee (S62817).

Blood samples were collected from 24 and 33 non-pregnant patients with SLE and CD, respectively. Additional 104 non-pregnant subjects without known diagnosis of cancer (from self-reported questionnaires) and normal cfDNA profiles were included. Written consent was obtained from all participants in study S60268 and S62795, and both studies were approved by the University Hospitals Leuven Ethics Committee.

Cell-free plasma DNA extraction and sequencing

Blood plasma was isolated through a standard two-step centrifugation procedure. cfDNA was extracted from 2 ml plasma using the Maxwell HT ccfDNA kit (Promega, Madison; automated procedure). DNA sequencing libraries of cfDNA were prepared using the KAPA HyperPrep kit (Roche Diagnostics). Whole-genome sequencing was carried out on a HiSeq2000, HiSeq2500, HiSeq4000, or NovaSeq sequencer (Illumina) generating 36 bp single-end or 51×2 paired-end reads. Raw reads were clipped to single-end 36 bp for standard processing and were mapped to the human reference genome GRCh38. Mean read count for downstream analysis was 9.96 million.

Sequencing data was subjected to the GIPSeq analysis pipeline as previously described¹⁶. In short, chromosomal z-scores are generated that measure the differential cfDNA representation, in standard deviations, between the sample and a reference set containing samples from women carrying a euploid fetus. A genome-wide quality score (QS), measuring the standard deviation of all autosomal z-scores following removal of the highest and lowest scoring chromosomes, was obtained. A NIPS result was marked as inconclusive when no reliable conclusion of fetal trisomy 21, 18 or 13 could be made, due to a low fetal fraction (<4%) or a deviating QS ($QS \geq 2$) in the GIPSeq profile. Cases with inconclusive NIPS results at the second sample were included in the study. Repeated inconclusive NIPS results that were classified as being suggestive of an underlying malignancy or a maternal constitutional copy number variation were excluded from the analysis. Among the selected inconclusive case series, two cases (one cfDNA sample from a pregnant woman diagnosed with SLE and one diagnosed with immune thrombocytopenic purpura) only had the first sample available. These two pregnant women both had repeated NIPS results in a previous pregnancy and refused a second sampling after the first inconclusive NIPS in the following pregnancy.

Smoothed bin counts (57,509 autosome bin features) from GIPSeq were further used for GIPXplore analysis. Principal component analysis (PCA) was used for dimension reduction and the top 50 principal components (PCs) were extracted for distance matrix construction using the Euclidean distance, followed by Walktrap community detection to define clusters with fixed parameters (the nearest number of nodes was 8 with a walk step of 2 and the subsequent optimal number of communities was determined by modularity). To visualize the dataset in lower dimensions, t-distributed stochastic neighbor embedding (tSNE) was used. For hierarchical clustering, the top 50 PCs and the Euclidean distances were used, and the hierarchical tree was constructed using Ward's linkage. PCA transformed genome-wide features were used for training in the machine learning model. Given that PCA was performed on the training data, the test data was projected onto the training PCA space for classification tasks. Performance was estimated by leave-one-out cross-validation and receiver operating characteristic (ROC) analysis. For the classifier, we used a support vector machine (SVM) and hyperparameters were chosen based on the grid search with 90% of the training data. Weighted sample size was accounted for in the model for imbalanced classes.

Statistical analysis

The proportion of conclusive and inconclusive (owing to a high QS) results were compared between pregnant women with and without AID using the odds ratio. The number of expected AID cases in the NIPS cohort was estimated using the population prevalence. Analogously, the odds ratio for pregnancy-induced IBD was computed. Outcomes of a high QS or a low FF among pregnant women with obesity and without obesity were compared using odds ratio as well. Logistic regression analysis was performed to determine an association between a given cluster and the proportion of AID cases. A *P* value was reported for clusters that were significantly enriched for cases with AID. A logit model was used to determine the log odds of the SLE prediction result in non-pregnant SLE cases with patient age, SLEDAI (Systemic Lupus Erythematosus Disease Activity Index) and disease activity as the independent variables. SLEDAI was categorized into four levels: not available, no activity (0), mild (1–5), and high (>6). Disease activity was categorized into two levels: yes (low, moderate, and high) and no.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

DATA AVAILABILITY

Processed alignments of sequencing data are archived to ArrayExpress (<https://www.ebi.ac.uk/arrayexpress/>) with unrestricted access under accession number E-MTAB-11607 and E-MTAB-10934. All other materials associated with this study are present in the paper or the Supplementary Information.

CODE AVAILABILITY

Code can be available upon reasonable request to the corresponding author.

Received: 14 January 2022; Accepted: 21 July 2022;

Published online: 14 September 2022

REFERENCES

1. Tan, E. M. & Kunkel, H. G. Characteristics of a Soluble Nuclear Antigen Precipitating with Sera of Patients with Systemic Lupus Erythematosus. *J. Immunol.* **96**, 464–471 (1966).

2. Abdelal, I. T., Zakaria, M. A., Sharaf, D. M. & Elakad, G. M. Levels of plasma cell-free DNA and its correlation with disease activity in rheumatoid arthritis and systemic lupus erythematosus patients. *Egypt. Rheumatol.* **38**, 295–300 (2016).
3. Koike, Y. et al. Dynamic pathology for circulating free DNA in a dextran sodium sulfate colitis mouse model. *Pediatr. Surg. Int.* **30**, 1199–1206 (2014).
4. Zablucova, Z. et al. Nuclear and Mitochondrial Circulating Cell-Free DNA Is Increased in Patients With Inflammatory Bowel Disease in Clinical Remission. *Front. Med.* **7**, 593316 (2020).
5. González, L. A., Ugarte-Gil, M. F. & Alarcón, G. S. Systemic lupus erythematosus: The search for the ideal biomarker. *Lupus* **30**, 181–203 (2021).
6. Verstockt, B. et al. Results of the Seventh Scientific Workshop of ECCO: Precision Medicine in IBD—Disease Outcome and Response to Therapy. *J. Crohns Colitis* **15**, 1431–1442 (2021).
7. Xu, Y. et al. High levels of circulating cell-free DNA are a biomarker of active SLE. *Eur. J. Clin. Invest.* **48**, e13015 (2018).
8. Truszcwska, A. et al. Cell-free DNA profiling in patients with lupus nephritis. *Lupus* **29**, 1759–1772 (2020).
9. Mondelo-Macia, P., Castro-Santos, P., Castillo-García, A., Muínelo-Romay, L. & Diaz-Peña, R. Circulating Free DNA and Its Emerging Role in Autoimmune Diseases. *J. Pers. Med.* **11**, 151 (2021).
10. Kubiritova, Z., Radvanszky, J. & Gardlik, R. Cell-Free Nucleic Acids and their Emerging Role in the Pathogenesis and Clinical Management of Inflammatory Bowel Disease. *Int. J. Mol. Sci.* **20**, 3662 (2019).
11. Duvvuri, B. & Lood, C. Cell-Free DNA as a Biomarker in Autoimmune Rheumatic Diseases. *Front. Immunol.* **10**, 502 (2019).
12. Bianchi, D. W. & Chiu, R. W. K. Sequencing of Circulating Cell-free DNA during Pregnancy. *N. Engl. J. Med.* **379**, 464–473 (2018).
13. Wan, J. C. M. et al. Liquid biopsies come of age: towards implementation of circulating tumour DNA. *Nat. Rev. Cancer* **17**, 223–238 (2017).
14. Vermeesch, J. R., Voet, T. & Devriendt, K. Prenatal and pre-implantation genetic diagnosis. *Nat. Rev. Genet.* **17**, 643–656 (2016).
15. Lo, Y. M. D. et al. Presence of fetal DNA in maternal plasma and serum. *Lancet* **350**, 485–487 (1997).
16. Bayindir, B. et al. Noninvasive prenatal testing using a novel analysis pipeline to screen for all autosomal fetal aneuploidies improves pregnancy management. *Eur. J. Hum. Genet.* **23**, 1286–1293 (2015).
17. Van Den Bogaert, K. et al. Outcome of publicly funded nationwide first-tier noninvasive prenatal screening. *Genet. Med* **23**, 1137–1142 (2021).
18. Lenaerts, L. et al. Comprehensive genome-wide analysis of routine non-invasive test data allows cancer prediction: A single-center retrospective analysis of over 85,000 pregnancies. *EClinicalMedicine* **35**, 100856 (2021).
19. van der Meij, K. R. M. et al. TRIDENT-2: National Implementation of Genome-Wide Non-Invasive Prenatal Testing as a First-Tier Screening Test in the Netherlands. *Am. J. Hum. Genet.* <https://doi.org/10.1016/j.ajhg.2019.10.005> (2019).
20. Mackie, F., Hemming, K., Allen, S., Morris, R. & Kilby, M. The accuracy of cell-free fetal DNA-based non-invasive prenatal testing in singleton pregnancies: a systematic review and bivariate meta-analysis. *BJOG Int. J. Obstet. Gynaecol.* **124**, 32–46 (2017).
21. Suzumori, N. et al. Classification of factors involved in nonreportable results of noninvasive prenatal testing (NIPT) and prediction of success rate of second NIPT. *Prenat. Diagn.* **39**, 100–106 (2019).
22. Hui, L. & Bianchi, D. W. Fetal fraction and noninvasive prenatal testing: What clinicians need to know. *Prenat. Diagn.* **40**, 155–163 (2020).
23. Samura, O. & Okamoto, A. Causes of aberrant non-invasive prenatal testing for aneuploidy: A systematic review. *Taiwan. J. Obstet. Gynecol.* **59**, 16–20 (2020).
24. MacKinnon, H. J. et al. The impact of maternal autoimmune disease on cell-free DNA test characteristics. *Am. J. Obstet. Gynecol. MFM* **3**, 100466 (2021).
25. Haghiaei, M. et al. Increased Death of Adipose Cells, a Path to Release Cell-Free DNA Into Systemic Circulation of Obese Women. *Obesity* **20**, 2213–2219 (2012).
26. Dabi, Y. et al. Autoimmune disorders but not heparin are associated with cell-free fetal DNA test failure. *J. Transl. Med.* **16**, 335 (2018).
27. Hui, C. Y. Y., Tan, W. C., Tan, E. L. & Tan, L. K. Repeated failed non-invasive prenatal testing in a woman with immune thrombocytopenia and antiphospholipid syndrome: lessons learnt. *Case Rep.* **2016**, bcr2016216593 (2016).
28. Hui, L., Bethune, M., Weeks, A., Kelley, J. & Hayes, L. Repeated failed non-invasive prenatal testing owing to low cell-free fetal DNA fraction and increased variance in a woman with severe autoimmune disease. *Ultrasound Obstet. Gynecol.* **44**, 242–243 (2014).
29. Che, H. et al. Pan-Cancer Detection and Typing by Mining Patterns in Large Genome-Wide Cell-Free DNA Sequencing Datasets. *Clin. Chem. hvac095* <https://doi.org/10.1093/clinchem/hvac095> (2022).
30. Rees, F., Doherty, M., Grainge, M. J., Lanyon, P. & Zhang, W. The worldwide incidence and prevalence of systemic lupus erythematosus: a systematic review of epidemiological studies. *Rheumatology* **56**, 1945–1961 (2017).
31. Rees, F. et al. The incidence and prevalence of systemic lupus erythematosus in the UK, 1999–2012. *Ann. Rheum. Dis.* **75**, 136–141 (2016).
32. Shand, A., Chen, J., Selby, W., Solomon, M. & Roberts, C. Inflammatory bowel disease in pregnancy: a population-based study of prevalence and pregnancy outcomes. *BJOG Int. J. Obstet. Gynaecol.* **123**, 1862–1870 (2016).
33. Kaul, A. et al. Systemic lupus erythematosus. *Nat. Rev. Dis. Prim.* **2**, 1–21 (2016).
34. Elkon, K. B. Review: Cell Death, Nucleic Acids, and Immunity. *Arthritis Rheumatol.* **70**, 805–816 (2018).
35. Heitzer, E., Auinger, L. & Speicher, M. R. Cell-Free DNA and Apoptosis: How Dead Cells Inform About the Living. *Trends Mol. Med.* **26**, 519–528 (2020).
36. Jiang, P. & Lo, Y. M. D. The Long and Short of Circulating Cell-Free DNA and the Ins and Outs of Molecular Diagnostics. *Trends Genet.* **32**, 360–371 (2016).
37. Goggs, R., Jeffery, U., LeVine, D. N. & Li, R. H. L. Neutrophil-Extracellular Traps, Cell-Free DNA, and Immunothrombosis in Companion Animals: A Review. *Vet. Pathol.* **57**, 6–23 (2020).
38. Aucamp, J., Bronkhorst, A. J., Badenhorst, C. P. S. & Pretorius, P. J. The diverse origins of circulating cell-free DNA in the human body: a critical re-evaluation of the literature. *Biol. Rev.* **93**, 1649–1683 (2018).
39. Papayannopoulos, V. Neutrophil extracellular traps in immunity and disease. *Nat. Rev. Immunol.* **18**, 134–147 (2018).
40. Lood, C. et al. Neutrophil extracellular traps enriched in oxidized mitochondrial DNA are interferogenic and contribute to lupus-like disease. *Nat. Med.* **22**, 146–153 (2016).
41. Chowdhury, C. S. et al. Elevated Levels of Total Cell-Free DNA in Maternal Serum Samples Arise from the Generation of Neutrophil Extracellular Traps. *Fetal Diagn. Ther.* **40**, 263–267 (2016).
42. Dinallo, V. et al. Neutrophil Extracellular Traps Sustain Inflammatory Signals in Ulcerative Colitis. *J. Crohns Colitis* **13**, 772–784 (2019).
43. Maronek, M. et al. Extracellular DNA Correlates with Intestinal Inflammation in Chemically Induced Colitis in Mice. *Cells* **10**, 81 (2021).
44. Wang, W.-M. & Jin, H.-Z. Role of Neutrophils in Psoriasis. *J. Immunol. Res.* **2020**, 3709749 (2020).
45. Anani, H. A. A. et al. Circulating Cell-Free DNA as Inflammatory Marker in Egyptian Psoriasis Patients. *Psoriasis Targets Ther.* **10**, 13–21 (2020).
46. Xiao, F. et al. NETosis may play a role in the pathogenesis of Hashimoto's thyroiditis. *Int. J. Clin. Exp. Pathol.* **11**, 537–547 (2018).
47. Fuchs, T. A. et al. Extracellular DNA traps promote thrombosis. *Proc. Natl Acad. Sci.* **107**, 15880–15885 (2010).
48. Chan, R. W. Y. et al. Plasma DNA aberrations in systemic lupus erythematosus revealed by genomic and methylomic sequencing. *Proc. Natl Acad. Sci.* **111**, E5302–E5311 (2014).
49. Malíčková, K. et al. Impaired Deoxyribonuclease I Activity in Patients with Inflammatory Bowel Diseases. *Autoimmune Dis.* **2011**, 945861 (2011).
50. Leffler, J. et al. A subset of patients with systemic lupus erythematosus fails to degrade DNA from multiple clinically relevant sources. *Arthritis Res. Ther.* **17**, 205 (2015).
51. Chan, R. W. Y. et al. Plasma DNA Profile Associated with DNASE1L3 Gene Mutations: Clinical Observations, Relationships to Nuclease Substrate Preference, and In Vivo Correction. *Am. J. Hum. Genet.* **107**, 882–894 (2020).
52. Lateef, A. & Petri, M. Managing lupus patients during pregnancy. *Best. Pract. Res. Clin. Rheumatol.* **27**, 435–447 (2013).
53. Knight, C. L. & Nelson-Piercy, C. Management of systemic lupus erythematosus during pregnancy: challenges and solutions. *Open Access Rheumatol. Res. Rev.* **9**, 37–53 (2017).
54. Zhang, C., Liang, M., Xu, X., Zhang, X. & Chen, S. Clinical features of new-onset systemic lupus erythematosus in pregnant patients. *J. Obstet. Gynaecol. Res.* **44**, 234–240 (2018).
55. Grömminger, S. et al. The influence of low molecular weight heparin medication on plasma DNA in pregnant women. *Prenat. Diagn.* **35**, 1155–1157 (2015).
56. Ma, G.-C., Wu, W.-J., Lee, M.-H., Lin, Y.-S. & Chen, M. Low-molecular-weight heparin associated with reduced fetal fraction and subsequent false-negative cell-free DNA test result for trisomy 21. *Ultrasound Obstet. Gynecol.* **51**, 276–277 (2018).
57. Torres, J. et al. Results of the Seventh Scientific Workshop of ECCO: Precision Medicine in IBD—Prediction and Prevention of Inflammatory Bowel Disease. *J. Crohns Colitis* **15**, 1443–1454 (2021).

ACKNOWLEDGEMENTS

We would like to thank the patients and blood donors. This study was supported by the Research Foundation-Flanders (FWO-Vlaanderen) (G080217N to JRV), Agentschap Innoveren en Ondernemen (VLAIO; Flanders Innovation & Entrepreneurship grant HBC.2018.2108 to TJ), EU H2020-MSCA-ITN-EJD-MATER (EU fund NUMBER 813707) to JRV and KS, EU H2020 research and innovation program under grant agreement No. 824110-EASI-Genomics to JRV, FWO-SBO (S003422) to JRV, KU Leuven funding (C14/18/092), and Clinical board for research and education Funding of the University

Hospitals Leuven mandate funding to LN and advanced project funding (RT 1065) to KVC. BV is supported by Clinical Research Fund (KOOR) from the University hospitals Leuven, Belgium.

AUTHOR CONTRIBUTIONS

H.C. and J.R.V. conceptualized and designed the study. H.C. and T.J. conducted project coordination. S.S., E.D.L. and B.V. carried out clinical sample procurement and processing. L.V., N.B., I.P. and K.V.D.B. analyzed NIPS profiles. L.L., K.D., S.S., E.D.L., S.V., B.V. and K.V.C. contributed to collection of clinical data. H.C. and L.D. performed bioinformatics analysis of sWGS data. H.C., T.J., L.L., K.S., L.D., E.D.L., B.V., K.V.C. and J.R.V. contributed to the interpretation of results. H.C., T.J. and J.R.V. wrote the manuscript; all co-authors reviewed the manuscript.

COMPETING INTERESTS

Patent application pending on 'Method for analyze cell-free nucleic acids' (JRV and LD). SV receive research grants from AbbVie, J&J, Galapagos, MSD, Pfizer, and Takeda; speaker fees from AbbVie, Falk, Ferring, Hospira, MSD, Pfizer, Takeda, and Tillotts; consultant for AbbVie, Avaxia, Celgene, Ferring, Galapagos, Genentech/Roche, Gilead, Roberts Clinical Trials, Hospira, Janssen, MSD, Mundipharma, Pfizer, Prodigest, Prometheus, Second Genome, Shire, and Takeda. BV is financially supported for research from Pfizer; lecture fees from AbbVie, Biogen, Chiesi, Falk, Ferring, Galapagos, Janssen, MSD, Pfizer, R-Biopharm, Takeda, and Truvion; consultancy fees from Applied Strategic, Atheneum, Bristol Myers Squibb, Guidepoint, Ipsos, Janssen, Progenity, Sandoz, Sosei Heptares, and Takeda.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41525-022-00325-w>.

Correspondence and requests for materials should be addressed to Joris Robert Vermeesch.

Reprints and permission information is available at <http://www.nature.com/reprints>

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022