



# Multi-Omic Approach to Identify Phenotypic Modifiers Underlying Cerebral Demyelination in X-Linked Adrenoleukodystrophy

Phillip A. Richmond<sup>1†</sup>, Frans van der Kloet<sup>2,3†</sup>, Frederic M. Vaz<sup>4,5</sup>, David Lin<sup>1</sup>, Anuli Uzozie<sup>6,7</sup>, Emma Graham<sup>1</sup>, Michael Kobor<sup>1</sup>, Sara Mostafavi<sup>1</sup>, Perry D. Moerland<sup>2</sup>, Philipp F. Lange<sup>6,7</sup>, Antoine H. C. van Kampen<sup>2,8</sup>, Wyeth W. Wasserman<sup>1‡</sup>, Marc Engelen<sup>9,10\*‡</sup>, Stephan Kemp<sup>4,5,9\*‡</sup> and Clara D. M. van Karnebeek<sup>1,11,12\*‡</sup>

## OPEN ACCESS

### Edited by:

Einat Zalcvar,  
Weizmann Institute of Science, Israel

### Reviewed by:

Myriam Baes,  
KU Leuven, Belgium  
Elaine L. Bearer,  
University of New Mexico Health  
Sciences Center, United States

### \*Correspondence:

Marc Engelen  
m.engelen@amsterdamumc.nl  
Stephan Kemp  
s.kemp@amsterdamumc.nl  
Clara D. M. van Karnebeek  
c.d.vankarnebeek@amsterdamumc.nl;  
c.d.vankarnebeek@amc.nl

<sup>†</sup>These authors share first authorship

<sup>‡</sup>These authors share last authorship

### Specialty section:

This article was submitted to  
Membrane Traffic,  
a section of the journal  
Frontiers in Cell and Developmental  
Biology

**Received:** 13 March 2020

**Accepted:** 02 June 2020

**Published:** 25 June 2020

### Citation:

Richmond PA, van der Kloet F, Vaz FM, Lin D, Uzozie A, Graham E, Kobor M, Mostafavi S, Moerland PD, Lange PF, van Kampen AHC, Wasserman WW, Engelen M, Kemp S and van Karnebeek CDM (2020) Multi-Omic Approach to Identify Phenotypic Modifiers Underlying Cerebral Demyelination in X-Linked Adrenoleukodystrophy. *Front. Cell Dev. Biol.* 8:520. doi: 10.3389/fcell.2020.00520

<sup>1</sup> Center for Molecular Medicine and Therapeutics, BC Children's Hospital Research Institute, University of British Columbia, Vancouver, BC, Canada, <sup>2</sup> Bioinformatics Laboratory, Department of Clinical Epidemiology, Biostatistics and Bioinformatics, Amsterdam Public Health Research Institute, Amsterdam University Medical Center, University of Amsterdam, Amsterdam, Netherlands, <sup>3</sup> Department of Pediatrics, Amsterdam University Medical Center, University of Amsterdam, Amsterdam, Netherlands, <sup>4</sup> Laboratory Genetic Metabolic Diseases, Amsterdam University Medical Center, University of Amsterdam, Amsterdam, Netherlands, <sup>5</sup> Department of Clinical Chemistry, Amsterdam Gastroenterology & Metabolism, Amsterdam, Netherlands, <sup>6</sup> Department of Pathology and Laboratory Medicine, University of British Columbia, Vancouver, BC, Canada, <sup>7</sup> Michael Cuccione Childhood Cancer Research Program, BC Children's Hospital, Vancouver, BC, Canada, <sup>8</sup> Biosystems Data Analysis, Swammerdam Institute for Life Sciences, University of Amsterdam, Amsterdam, Netherlands, <sup>9</sup> Department of Pediatric Neurology, Amsterdam Neuroscience, Amsterdam Leukodystrophy Center, Emma Children's Hospital, Amsterdam University Medical Center, University of Amsterdam, Amsterdam, Netherlands, <sup>10</sup> Department of Neurology, Amsterdam University Medical Center, University of Amsterdam, Amsterdam, Netherlands, <sup>11</sup> Department of Pediatrics, Emma Children's Hospital, Amsterdam University Medical Center, University of Amsterdam, Amsterdam, Netherlands, <sup>12</sup> Department of Pediatrics, Amalia Children's Hospital, Radboud University Medical Center, Nijmegen, Netherlands

X-linked adrenoleukodystrophy (ALD) is a peroxisomal metabolic disorder with a highly complex clinical presentation. ALD is caused by mutations in the *ABCD1* gene, and is characterized by the accumulation of very long-chain fatty acids in plasma and tissues. Disease-causing mutations are 'loss of function' mutations, with no prognostic value with respect to the clinical outcome of an individual. All male patients with ALD develop spinal cord disease and a peripheral neuropathy in adulthood, although age of onset is highly variable. However, the lifetime prevalence to develop progressive white matter lesions, termed cerebral ALD (CALD), is only about 60%. Early identification of transition to CALD is critical since it can be halted by allogeneic hematopoietic stem cell therapy only in an early stage. The primary goal of this study is to identify molecular markers which may be prognostic of cerebral demyelination from a simple blood sample, with the hope that blood-based assays can replace the current protocols for diagnosis. We collected six well-characterized brother pairs affected by ALD and discordant for the presence of CALD and performed multi-omic profiling of blood samples including genome, epigenome, transcriptome, metabolome/lipidome, and proteome profiling. In our analysis we identify discordant genomic alleles present across all families as well as differentially abundant molecular features across the omics technologies. The analysis was focused on univariate modeling to discriminate the two phenotypic groups, but was unable to identify statistically significant candidate molecular markers. Our study

highlights the issues caused by a large amount of inter-individual variation, and supports the emerging hypothesis that cerebral demyelination is a complex mix of environmental factors and/or heterogeneous genomic alleles. We confirm previous observations about the role of immune response, specifically auto-immunity and the potential role of PFN1 protein overabundance in CALD in a subset of the families. We envision our methodology as well as dataset has utility to the field for reproducing previous or enabling future modifier investigations.

**Keywords:** multi-omics, LCMS, methylation, lipidomics, proteomics, bioinformatics, X-ALD, cerebral demyelination

## INTRODUCTION

Adrenoleukodystrophy (ALD) is a rare peroxisomal X-linked degenerative disease (MIM 300100), caused by deficiency of the ABC half-transporter encoded by the *ABCD1* gene. Over 850 different disease-causing loss-of-function *ABCD1* mutations have been reported<sup>1</sup>. Mutations lead to a defect in the import of very long-chain fatty acids (VLCFA) into peroxisomes for further degradation and a subsequent accumulation of VLCFA in plasma and tissues. The overall incidence is 1:15,000 (Moser et al., 2016). In males, ALD often manifests with adrenocortical insufficiency in childhood (50% before 10 years) (Huffnagel et al., 2019b). During adulthood virtually all male and, eventually, female patients develop a progressive myelopathy termed adrenomyeloneuropathy (Engelen et al., 2012; Engelen et al., 2014; Huffnagel et al., 2019a). Additionally, during childhood or sometimes through adulthood male patients can develop cerebral demyelination, termed cerebral ALD (CALD). It is estimated that eventually more than 60% of male patients develop CALD (Kemp et al., 2012; de Beer et al., 2014). Untreated CALD is often progressive, but can spontaneously arrest in 10–20% of patients. It causes vegetative state and death 2–3 years after onset, so early identification as well as careful and frequent monitoring of all male ALD patients is necessary. If diagnosed early, hematopoietic stem cell therapy can be used to halt further progression of cerebral ALD (Pierpont et al., 2017). However, the precise mechanism by which hematopoietic stem cell therapy arrests CALD progression is not clear. To ensure timely stem cell therapy for males with CALD, affected individuals are subjected to rigorous neurological and MRI follow-ups that pose considerable physical, emotional and financial burden. As such, the unresolved and unpredictable phenotypic variability of ALD is a crucial roadblock for patient care.

As newborn screening for ALD has recently been implemented, there is an urgent need for identification of markers which may be prognostic of cerebral demyelination in many newly diagnosed patients around the world (Moser et al., 2016; Barendsen et al., 2020). Our research focuses on delineating the enormous phenotypic variability in ALD, with the overarching goal of identifying biomarkers prognostic of the advancement to CALD. If successful in identifying biomarkers with prognostic power, then the biomarkers could

replace existing expensive monitoring protocols and potentially highlight therapeutic targets, as is the case with other rare genetic disorders. For example, in Spinal Muscular Atrophy, the genes *Plastin-3 (PLS3)* and *Coronin 1C (CORO1C)* were identified as protective modifiers, unraveling impaired endocytosis as a rescue mechanism for the phenotype (Hosseinibarkooie et al., 2016). These modifiers were identified from studies focusing on siblings with discordant disease severity, and are opening novel therapeutic targets for treatment. Patients with ALD may benefit from similar research advances.

Phenotypic discordance in individuals with the same *ABCD1* genotype, including siblings and even monozygotic twins (Korenke et al., 1996), strongly supports the hypothesis that other modifying factors play a role in the progression of the disease (Wiesinger et al., 2015; Kemp et al., 2016). As yet, however, modifier studies using candidate gene approaches have had little success and resulted in the identification of only a single modifier gene, cytochrome P450 family 4 subfamily F member 2 (*CYP4F2*), with limited prognostic power (van Engen et al., 2016). Other candidate variants have been proposed, including a candidate *cis*-regulatory SNP in the promoter region of *ELOVL1* fatty acid elongase 1 (*ELOVL1*)—a gene involved in VLCFA synthesis (Ofman et al., 2010). The functional consequences of this SNP with respect to the expression of *ELOVL1* in the brain is still under investigation (Kemp et al., 2012). The lack of modifier identification could be due to the limited genomic search space that was explored, which to date has focused only on candidate gene approaches. Owing to the small sample size inherent to rare disease cohorts, traditional genome wide association studies (GWAS) approaches are not feasible. Employing a strategy which utilizes family structure may allow for a narrower search space compared to GWAS, while allowing a broader interrogation of the genome than candidate gene approaches. Beyond a search space which involves genetic mapping, other high throughput “omics” technologies allow the exploration of complex biological systems at many levels. It is now possible to identify differences between individuals or phenotypic states at the DNA, methylated DNA, RNA, lipid, and protein levels. Our goal is to delineate personal molecular characteristics that contribute to phenotypic variability in male ALD siblings enabling the identification of biomarkers that prognosticate onset and progression of CALD. Because modifying factors could also include environmental, epigenetic and microbiome factors (Génin et al., 2008; Argmann et al., 2016), multi-omics approaches are key.

<sup>1</sup>www.adrenoleukodystrophy.info

**TABLE 1** | Summary of patients within ALD cohort.

Family	Age	Phenotype	<i>ABCD1</i> mutation	<i>ELOVL1</i> (A > G)	<i>CYP4F2</i> (C > T)	<i>APOE</i> rs429358	<i>APOE</i> rs7412	<i>APOE</i> Genotype
1	28	CALD	c.1390C>T	G/G	T/T	T/T	C/C	ε3/ε3
1	28	Non-CALD		G/G	C/T	T/T	C/C	ε3/ε3
2	30	CALD	c.1899delC	A/G	C/C	T/T	C/C	ε3/ε3
2	30	Non-CALD		A/G	C/C	T/T	C/C	ε3/ε3
3	38	CALD	c.1992-2A>G	A/G	C/C	T/T	C/C	ε3/ε3
3	36	Non-CALD		A/G	C/C	T/T	C/C	ε3/ε3
4	6	CALD	c.659T>C	A/A	C/T	T/T	T/T	ε2/ε2
4	8	Non-CALD		A/A	C/T	T/T	T/T	ε2/ε2
5	16	CALD	c.1866-2A>T	A/G	C/C	T/C	C/C	ε3/ε4
5	18	Non-CALD		G/G	C/C	T/C	C/C	ε3/ε4
6	27	CALD	c.892G>A	A/A	C/C	T/C	C/C	ε3/ε4
6	25	Non-CALD		A/G	C/C	T/C	C/C	ε3/ε4

The family number, age at sample collection, ALD phenotype, *ABCD1* variant, and genotypes for previously associated modifier alleles for all patients within the cohort.

In this study, we carefully selected a set of six well characterized brother pairs who have the same *ABCD1* pathogenic allele but are discordant for cerebral ALD: one brother has CALD and the other has no white matter lesions on MRI (non-CALD). The brother pairs are close in age (no more than 2 years apart), and range in age from 6–38 years at sample collection (Table 1). Blood samples were obtained from each of these patients and underwent profiling through five omics technologies (Figure 1) including whole genome sequencing (WGS), RNA sequencing (RNA-seq), EPIC DNA methylation (DNAm) microarray, lipidomic profiling via liquid chromatography mass spectrometry (LCMS), and protein profiling by LCMS. Within this study our choice to use blood samples, either lymphocytes or plasma, is inherent to our research question which is whether we can identify molecular markers of cerebral demyelination in an accessible sample in patients. Uncovering the pathophysiology of cerebral demyelination would require sampling of tissue from the brain and the blood–brain barrier, which is beyond the scope of this project. Each omics dataset was processed to quantify/map features, undergo quality control analysis, and then used for group-wise comparisons between CALD and non-CALD phenotype groups using univariate analysis. We first investigated the potential for a single, shared modifier allele which could discriminate the two groups from the WGS data. Next, we systematically compared the groups for each of these omics data sets to find potential markers specific to the phenotype. We aggregated the datasets together after performing pairwise comparisons and identified heterogeneous signals within sub-groups of the six families. To the best of our knowledge this is the most comprehensive study to date in terms of systems biology characterization of human ALD using a unique collection of samples.

## MATERIALS AND METHODS

### Project Overview

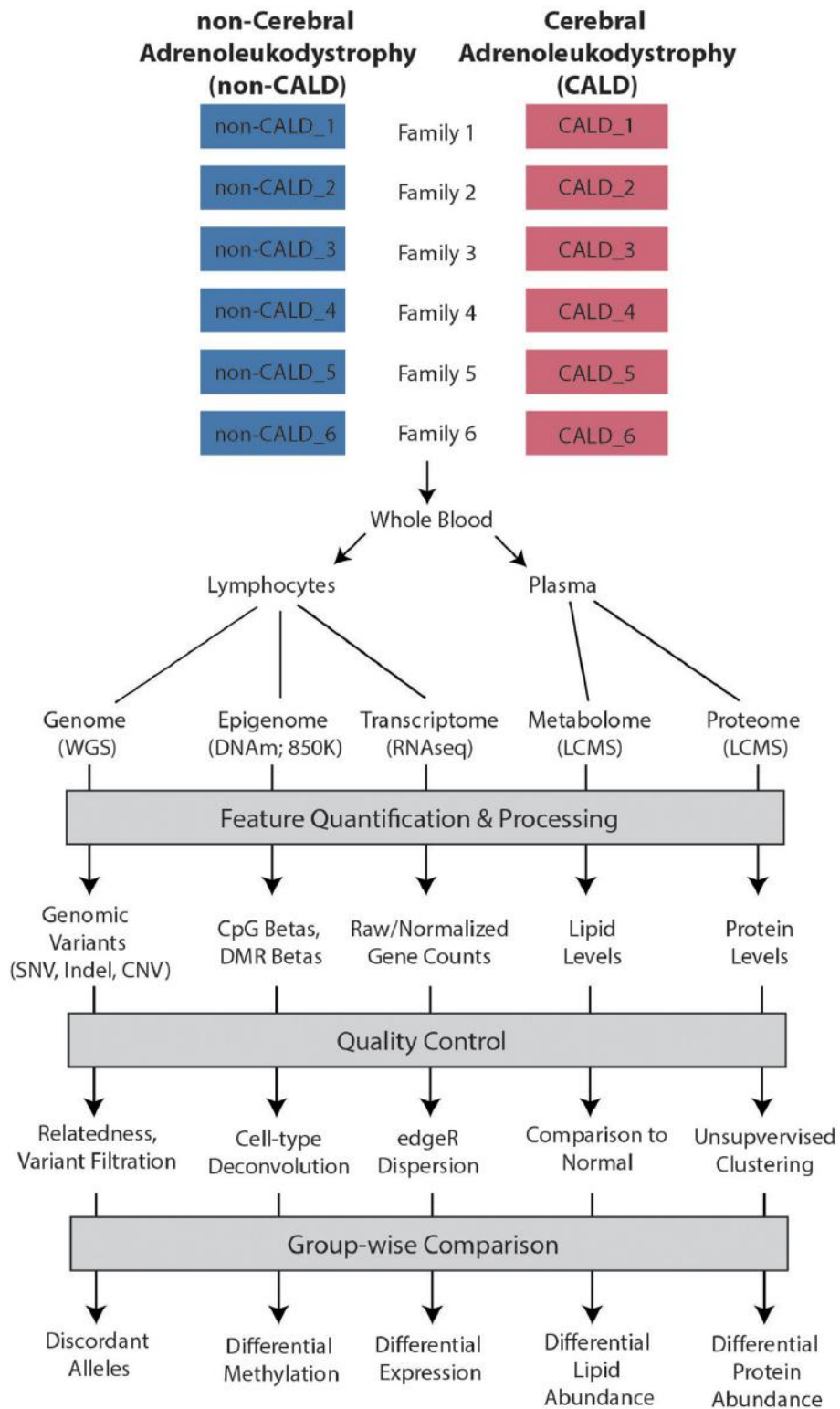
An overview of the project can be found in Figure 1, which depicts the project phases including patient

phenotyping/sample collection, multi-omic data collection, feature quantification/processing, quality control, and group-wise comparisons between phenotype groups. In this project, six brother pairs affected by ALD but discordant for the presence of cerebral ALD were included. We only selected the brother pairs for analysis, and did not include siblings or parents as our focus was on differences between two patients both affected by *ABCD1* mutations. Patients were selected from the Dutch cohort, an ongoing prospective natural history study (Huffnagel et al., 2019c). Blood was drawn from the brother pairs and lymphocyte pellets or plasma was isolated. Lymphocyte pellets were used for whole WGS, RNA-seq, and DNAm. Fasted plasma was used for downstream LCMS analysis identifying either lipid or protein abundances. For control samples, anonymized adult male plasma samples were used in the lipidomics analysis. Data was then processed independently for each of the platforms including feature quantification/mapping, followed by platform specific quality control and group-wise comparisons. Details regarding sample collection, platform specifications, and specific methodology for each analysis performed in this project can be found in the **Supplementary Methods** section.

### Patient Selection and Phenotyping

All patients were selected from the Dutch cohort, an ongoing prospective cohort study. All patients are examined yearly (by ME) and undergo an MRI of the brain at the time of examination. Samples are collected in the PEROX biobank. The presence of cerebral ALD is defined as the presence of white matter lesions in a distribution consistent with ALD. The classification of the sibs (CALD versus non-CALD) is valid at this time, but the non-affected individuals can theoretically convert to cerebral ALD.

All samples were collected and stored in the PEROX Biobank according to a protocol (METC2015\_066) approved by the biobank review board of the Amsterdam UMC (BioBank Toetsingscommissie AMC). All patients provided written informed consent for storage and use of materials for medical research.



**FIGURE 1 |** Project overview. An overview of the data and processes involved in the project including samples from two brothers across unrelated families, blood isolated into lymphocytes and plasma, and then profiling with five omics technologies including WGS for the genome, DNA methylation (DNAm) via the 850K EPIC microarray, transcriptome profiling with RNA sequencing (RNA-seq), metabolome profiling with liquid chromatography mass spectrometry (LCMS), and protein profiling with LCMS. These data are then taken through feature quantification/processing, quality control metrics, and group-wise comparison through univariate modeling.



## Feature Quantification and Data Processing

For each platform, the data was processed independently following best-practices guidelines from the groups generating the datasets. Details regarding feature quantification and assignment at the gene, lipid, protein, and differentially methylated region (DMR) level can be found in the **Supplementary Methods** section.

## Univariate Modeling of CALD vs. Non-cerebral ALD

Using univariate modeling techniques the prognostic power of each lipid, transcript (RNA) or protein is calculated as:

$$y = \rho\beta + fam + \varepsilon \quad (1)$$

in which  $y$  is the observed value and  $\rho$  the phenotype (0 or 1),  $\beta$  is the weight and  $fam$  the family cofactor.  $\varepsilon$  is the remaining error. Because methylation of DNA changes with *age* (McEwen et al., 2018), *age* is also included as a cofactor:

$$y = \rho\beta + fam + age + \varepsilon \quad (2)$$

when analyzing the DNAm results. The significance ( $p$ -value) of the discriminating phenotype (fixed) and family (random) effects are determined by ordinary least squares modeling (OLS) of the data using the model from Eq. 1 in case of lipid and proteomic data (Harrison et al., 2018). In the case of methylation and RNA sequencing data the  $p$ -values are determined by maximum likelihood estimates (MLE) of the fixed and random effects using Limma (Ritchie et al., 2015) and edgeR, respectively (Robinson et al., 2010; Ritchie et al., 2015).

## Allele Comparisons in Whole Genome Sequencing Data

Details on data processing, including variant calling and comparing across samples can be found in the **Supplementary Methods** section. Briefly, allele comparisons were performed in whole genome sequencing data on jointly genotyped variant datasets. For SNVs and indels, variants were jointly genotyped and converted into a GEMINI database (Paila et al., 2013). This database was then queried to identify subsets of discriminating alleles. For structural variants and mobile element insertions, custom scripts were used to identify discordant genotypes from annotated jointly genotyped variant tables. Discordant genotypes, stored as unique variant identifiers, were then placed into Intervene for intersection analysis (Khan and Mathelier, 2017).

## Aggregation of Signal Across Platforms

To assess the added value of combining the different platforms, significant signals prior to multiple testing correction were collected for each omics platform and intersected at the annotated gene level (hg19). Because of the lack of a clear mapping of lipids to genes the lipidomics platform was excluded from this intersection allowing 4 possible intersections; DNAm-RNA, DNAm-Protein, RNA-Protein, and the overall intersection

of DNAm-RNA-Protein. Further investigation of a shared signal was performed by clustering the first three principal components (i.e., capturing the most variance) of the log fold changes (top 10 and  $p < 0.05$ ) of the combined platform data (including lipids).

## Assessing Contribution of Family Effect per Feature

The data were modeled using the equations (above) in which both phenotypic and family effects are estimated. We partitioned the variance for the lipids, proteins, and RNA datasets to identify the contribution of the family effect, the phenotype effect, or the residual variance using the variancePartition package (Hoffman and Schadt, 2016). The same was repeated for the DNA methylation dataset with the addition of the *age*, and *phenotype:age* variance terms. Next, we plotted the top two principal components for each omics dataset before and after the removal of the variance contributed from the family effect with the limma:removeBatchEffect tool (Ritchie et al., 2015). Lastly, to determine the sensitivity/specificity of the findings for leaving out a one or two families all the analyses (excluding methylation data analysis) that were run for the case of all families were repeated with a one or two families left out (e.g., without fam 1, without fam 2, without fam 1 and fam 2, etc.). We encapsulated this information in separate upset plots for each platform.

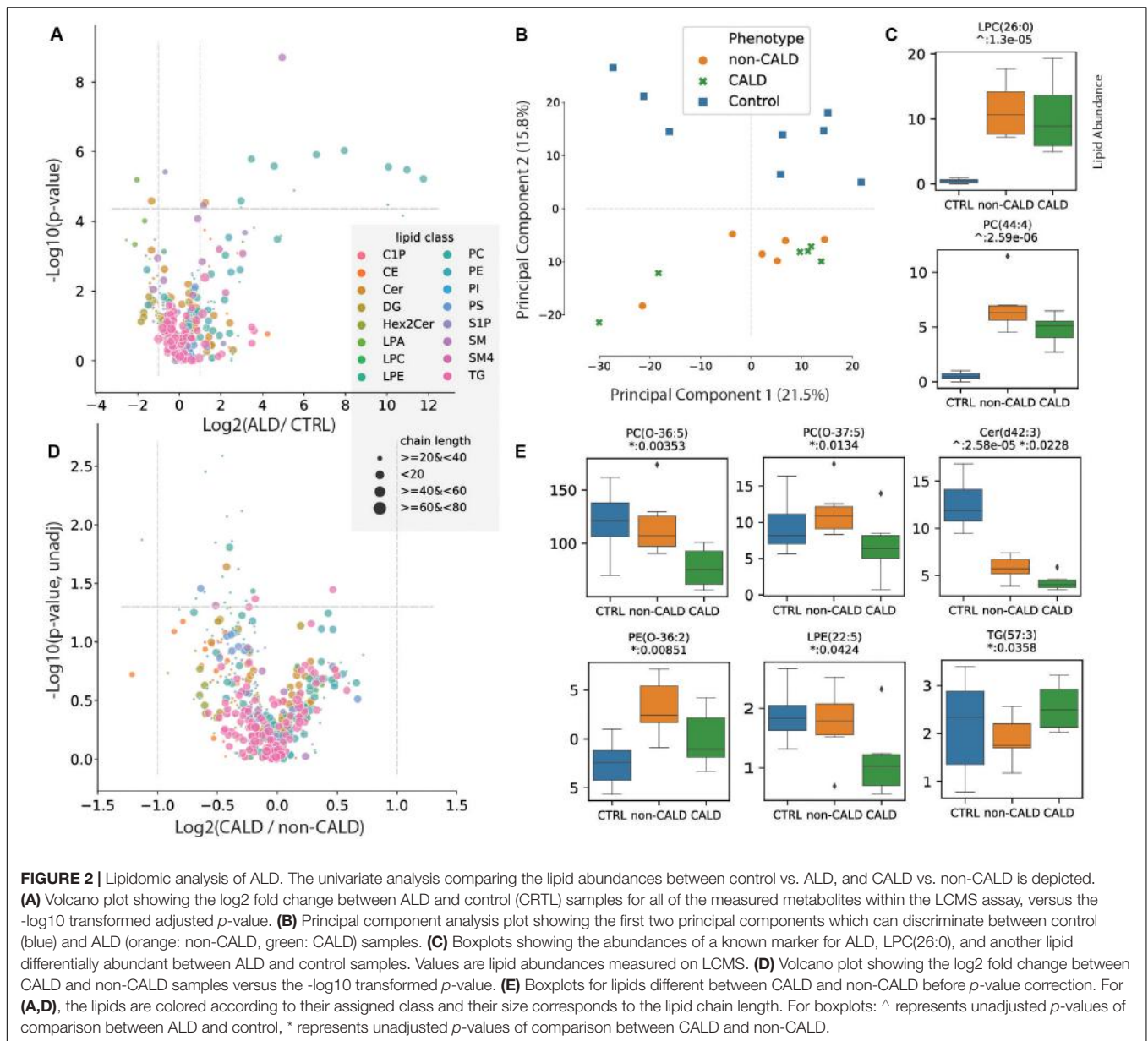
## RESULTS

### Lipidomics Analysis of a Fatty Acid Storage Disorder

Patients affected by ALD have a buildup of VLCFAs within cells in the body. Recent mass spectrometry advances allow for broad, untargeted profiling of lipids (Huffnagel et al., 2019a). We applied LCMS from plasma samples of each of the patients within this cohort as well as matched control samples (eight fasted plasma samples from healthy males were used as controls).

First, we identified differential lipid abundances between ALD (both non-CALD and CALD) samples and control samples, with 139 lipids passing the threshold of  $p$ -value  $< 0.05$  (Eq. 1, OLS), and 17 lipids remaining significant after multiple testing correction (Bonferroni) (**Supplementary Figure S1** and **Supplementary Table S1**) (Methods). The measured lipids are plotted as a volcano plot, that is the log<sub>2</sub> fold change of ALD over control versus corrected  $p$ -value (**Figure 2A**). We confirm that untargeted lipidomic profiling can distinguish ALD from control samples via principal component analysis, and also capture the expected differentially abundant lipids between control and ALD samples including the known ALD biomarker LPC(26:0) (**Figures 2B,C**).

Next, we compared CALD and non-CALD groups for differences in lipid abundance which could act as markers of cerebral demyelination. Of note, the principal component analysis which separates ALD from control did not separate CALD from non-CALD, i.e., the differences in lipid profiles between these two phenotypes are much less pronounced than the differences separating ALD patients from controls



(Figure 2B). The measured lipids are plotted as a volcano plot, that is the log<sub>2</sub> fold change of CALD over non-CALD versus transformed *p*-value (Figure 2D). In total 22 lipids were found to have different abundances between the two groups with *p*-value < 0.05, however, none of the lipids remained significant after correcting for multiple testing (Supplementary Table S2). The observed differences are much smaller between CALD and non-CALD compared to ALD and control, as highlighted by the differences in fold change axes (Figures 2A,D). Interestingly, there was a higher abundance in the non-CALD group for several key VLCFAs involved in ALD including PC(44:4) and Cer(d42:3), the latter reaching *p*-value < 0.05 (Figures 2C,E). While some lipids show a relatively large fold change between CALD and non-CALD groups as a whole, the signal is not consistent for every family. An example of this can be seen

in SM(d36:2) or PS(43:3) (Supplementary Figure S2). This limits the prognostic power of these lipids as consistent markers delineating the phenotype. Lastly, we observed a large range of lipid abundances within the control group for several of the differential lipids between CALD and non-CALD, which could indicate that these lipids are variable within healthy individuals and the signal we observe between CALD and non-CALD could be due to noise or variation in the healthy population (Figure 2E and Supplementary Figure S1).

### Discordant Genotype Analysis for the Identification of a Modifier Allele

Using whole genome sequencing, we investigated a range of variant classes for discordant alleles between siblings. These

discordant alleles are then intersected across multiple families under the hypothesis that polymorphic differences contribute to cerebral demyelination.

We first focused on alleles which emerged from previous modifier studies to see if they are confirmed. Proposed modifier alleles from target gene studies have identified two candidates within *ELOVL1* (rs839765) and *CYP4F2* (rs2108622) (Kemp et al., 2012; van Engen et al., 2016). Within this cohort, those modifier alleles do not segregate with ALD phenotype (Table 1), nor are the genotypes shared or lacking in the confidently phenotyped CALD patients. Furthermore, it has been suggested that apolipoprotein E (*APOE*) genotypes—which are a combination between two SNP sites to produce *APOE2* ( $\epsilon 2$ ), *APOE3* ( $\epsilon 3$ ), and *APOE4* ( $\epsilon 4$ ) alleles—may be markers of disease severity and cerebral progression (Orchard et al., 2019a). These *APOE* alleles do not segregate with disease nor are they shared by all CALD patients. Together, these results suggest limited prognostic power of these alleles, and perhaps supports heterogeneous contributions of genetic background to disease progression.

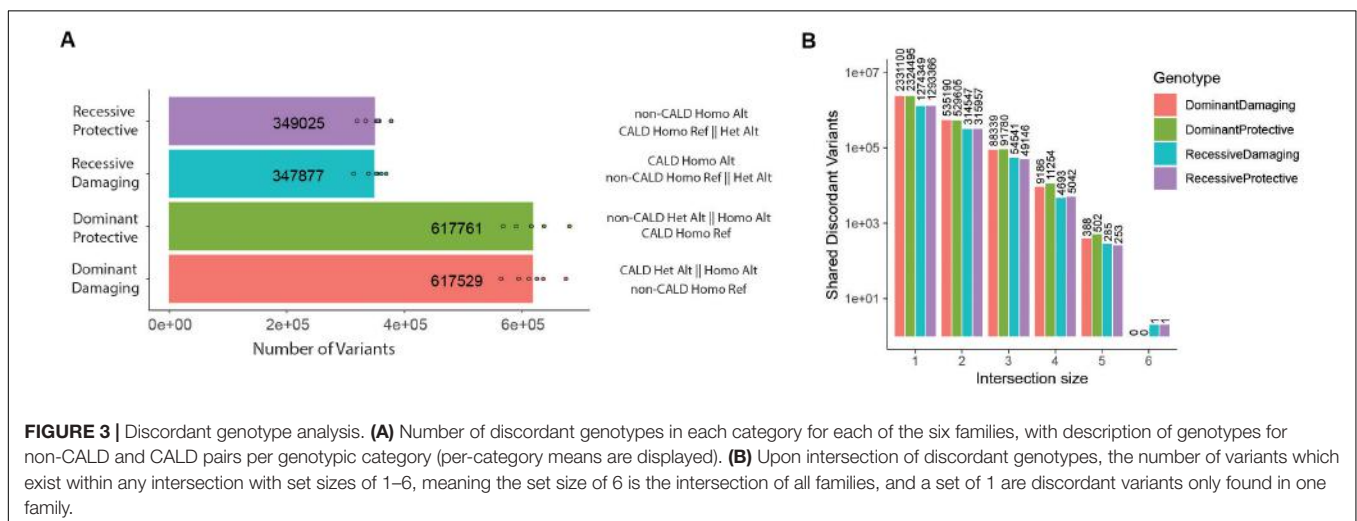
Next, for several variant classes, we performed a discordant analysis between siblings and intersected these alleles across families (Figure 3). We considered four genotypic categories termed dominant protective, recessive protective, dominant damaging, or recessive damaging based on the genotype (heterozygous: dominant or homozygous: recessive) and the sibling which carries the genotype (CALD: damaging or non-CALD: protective). Performing this genotypic analysis on SNVs and indels, we identified  $\sim 6.0 \times 10^5$  discordant candidate variants in the dominant categories from each family, and  $\sim 3.0 \times 10^5$  discordant candidate variants from the recessive categories (Figure 3A and Supplementary Table S3). Despite the large number of discordant candidates per family, intersecting these sets across families reduces the candidates dramatically, resulting in only two candidate variants at the intersection of all six families (Figure 3B) (Supplementary Table S4). A recessive damaging variant downstream of the PYM homolog 1 (*WIBG/PYM1*) gene (rs7980776) and recessive protective

allele (rs55639747/rs61327784) within the intronic region of the deuterosome assembly protein 1 (*CCDC67/DEUP1*) gene (Supplementary Figures S3, S4). We validated our approach with a parallel pipeline utilizing the new DeepVariant tool (Poplin et al., 2018), which claims higher accuracy than GATK HaplotypeCaller (Supplementary Table S4). There is high concordance between the two variant call sets, and they produced the same two variants within the intersection. A single additional variant was reported using DeepVariant under the recessive damaging model, however, the variant did not pass the manual inspection quality assessment. *In silico* analysis of both variants suggests these variants have little functional effect, and the associated genes did not link to the cerebral demyelination phenotype (Supplementary Results).

For the other variant classes, including structural variants (SVs) and mobile element insertions (MEIs), we performed joint genotyping to identify shared and discordant alleles in the same manner as SNVs and indels. We identified  $\sim 1,500$  and  $\sim 400$  SVs in the dominant and recessive categories, respectively, and  $\sim 400$  and  $\sim 100$  MEIs (Supplementary Table S3). Unsurprisingly, these discordant events were not shared across more than 4 families (Supplementary Table S4). We further manually inspected the regions around the discordant SNVs/indels identified above, and did not find any other segregating SVs or MEIs.

Lastly, we extended our discordance analysis to the mitochondrial genome to examine candidate alleles which may show evidence of heteroplasmy which are not shared between two siblings. We identified that between 129 and 547 mitochondrial variants per sample, of which 52 to 476 are heteroplasmic, and none are consistent discriminating variants between phenotypes shared across all families (Supplementary Table S5). Further, if we aggregated at the gene level, we did not find any heteroplasmic variants consistent across the same gene.

In recognition of a study limitation – the fact that some non-CALD patients may progress to CALD – we further intersected alleles shared by all CALD patients. These variants were annotated by impact or as eQTLs defined in GTEx





(**Supplementary Methods**). There were 48 variants present in the heterozygous state across all CALD patients, where no non-CALD patients were heterozygous (**Supplementary Table S6**). Of these, a haploblock containing 20 variants was identified overlapping the two pore segment channel 2 (*TPCN2*) gene, including a missense variant (rs3750965) (**Supplementary Figure S5**). Interestingly, the only patient which was homozygous for this variant is the youngest non-CALD patient within the cohort, suggesting that this gene could be of significance should the patient develop the cerebral demyelination phenotype.

## Univariate Modeling of Phenotype Differences Across Omics Platforms

Beyond identifying a single genetic modifier allele, the omics platforms allow for the identification of candidate molecular signatures which can discriminate between the CALD and non-CALD phenotypes. Using univariate analysis, we identify differences across each platform at the feature level, to search for a signal which can be used as a marker for transition to CALD. Further, we leverage these molecular signatures to provide insight into the pathogenesis of cerebral demyelination.

### Transcriptomics

Examining RNA expression using RNA-seq provides a measurement for nearly all expressed protein coding genes in the genome. Differential gene expression was calculated between the two phenotype groups using the univariate model accounting for family effect (Eq. 1). There were 199 genes found with a  $p$ -value  $< 0.05$ , although none remained significant after multiple testing correction (Bonferroni) (**Figures 4A,B** and **Supplementary Figure S6**). This is likely due to the low number of samples and relatively small differences that were observed between the two groups. Furthermore, many of the genes identified as significant were inconsistent in one or more of the sibling pairs, limiting the diagnostic utility as a marker (**Figure 4B**). Despite not having significant genes after multiple testing correction, we performed enrichment analysis using GO (gene annotation) and KEGG (pathway annotation) to derive insights based on the 199 genes passing a threshold of  $p$ -value  $< 0.05$  (**Supplementary Figure S7**). Of note, elevated interferon related processes suggest that the host may be reacting to pathogens activating the immune system (Hoffmann et al., 2015). It is therefore no surprise that 3 chemokines (*CXCL6*, *CXCL8* and *IFI27*) were found in the top 10 differentially expressed genes. Amongst the remainder of the proteins encoded by the top 10 differentially expressed genes, the D-Xylulokinase gene (*XYLB*) encodes for the protein that catalyzes the ATP-dependent phosphorylation of D-xylulose to produce xylulose-5-phosphate (Xu5P) therefore *XYLB* may play an important role in metabolic disease given that Xu5P is a key regulator of glucose metabolism and lipogenesis (Bunker et al., 2013). The glycine amidinotransferase (*GATM*) gene has been associated with statin intolerance (Willrich et al., 2018) and its function to catalyze creatine and possibly affect the production of ceramides (Turer et al., 2017). Lastly, the myosin-binding protein 1 (*MYO1B*) gene codes for a protein that may participate in a process critical to neuronal development and function such

as cell migration, neurite outgrowth and vesicular transport (Sittaramane and Chandrasekhar, 2008).

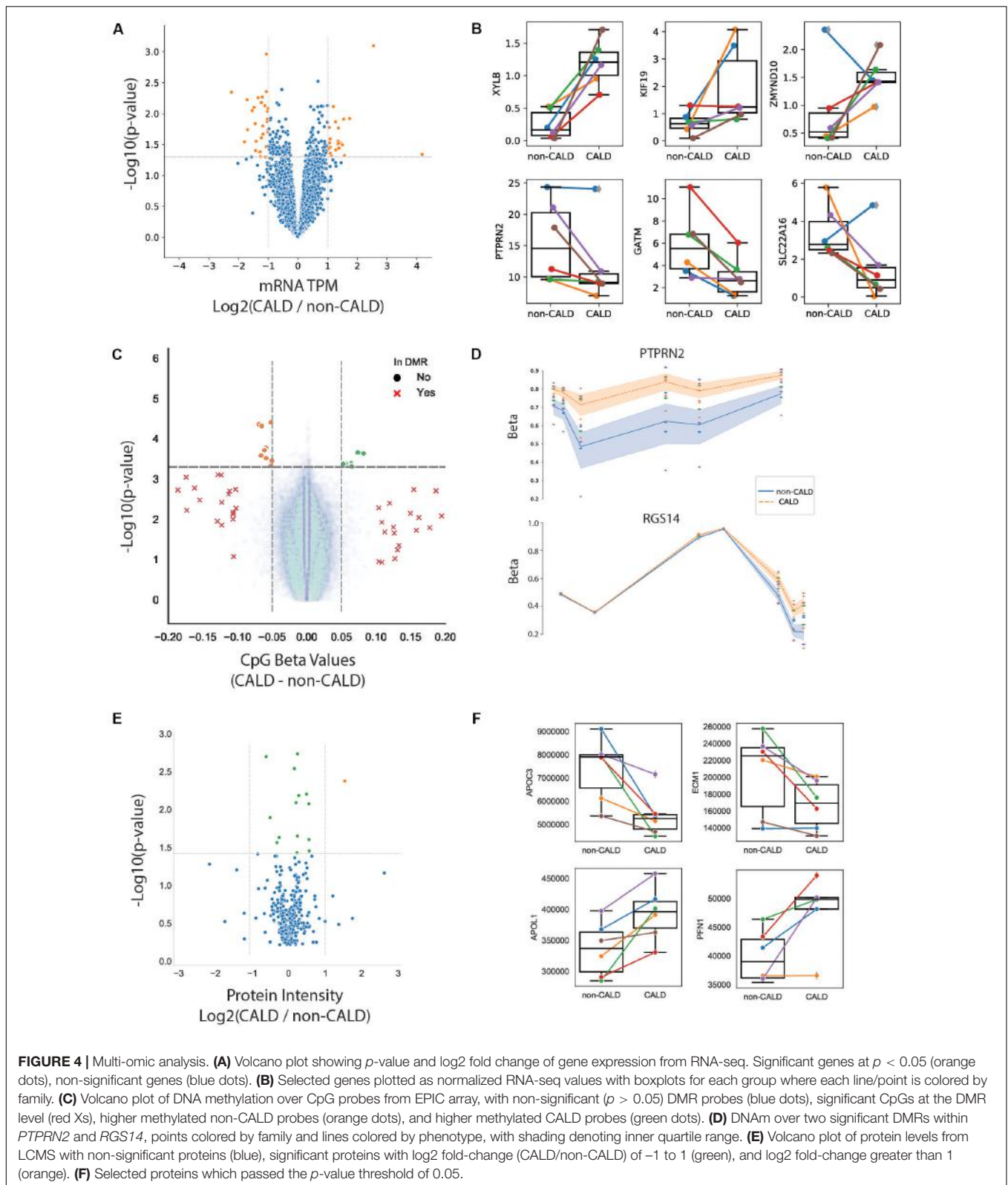
### Epigenomics

DNA methylation has been linked to changes in gene expression, and is an important readout of some environmental impacts upon the cell. Measuring DNA methylation is typically done at specific methylation sites (CpGs), and then aggregated across regions where several sites have similar trends of methylation levels to find differentially methylated regions (DMRs). Here, we used the MethylationEPIC BeadChip which targets over 850,000 CpGs. Using LIMMA modeling including age as cofactor (Eq. 2) 264 CpGs had a nominal  $p$ -value  $< 0.0005$ . Of these 264 CpGs, 16 passed the delta beta (i.e., difference between methylation levels of CALD vs. non-CALD) of  $> 5\%$  (**Supplementary Table S8**). When aggregating these loci into a DMR analysis, we identified 22 regions passing thresholds of FDR  $< 0.05$  and  $> 10\%$  methylation change (**Figure 4C**). Multiple CpGs map to the same gene and show a large delta beta, which we identified in the genes protein tyrosine phosphatase receptor type N2 (*PTPRN2*) and regulatory of G protein signaling 14 (*RGS14*) (**Figure 4D**). *RGS14* may alter calcium levels to enhance long term potentiation and learning (Lee et al., 2010). Due to its presence in neurosecretory vesicles, *PTPRN2* has been implicated in insulin and neurotransmitter exocytosis (Sengelaub et al., 2016). Furthermore, *PTPRN2* hypermethylation has been identified within a separate study which compared DNA methylation between CALD and non-CALD patients (Schlüter et al., 2018).

### Proteomics

In addition to profiling lipids, LCMS can be used for high throughput profiling of proteins thus enabling the identification of differential protein abundances between samples. Applying proteomics to these 12 patients yielded a quantification of 5,862 peptides which were matched against 351 protein groups. Comparing CALD and non-CALD groups, we found 16 proteins with differential abundances ( $p < 0.05$ ) (**Figures 4E,F**, **Supplementary Figure S8**, and **Supplementary Table S9**). Investigating the top hits we find 4/16 proteins associated with immunoglobulin heavy chain (IGHV4-34, IGHV3-30, IGHV3-7, and P0DOX6), 2/16 are associated with immunoglobulin kappa variables (IGKV6D-21 and IGKV1D-33), and with P0DOX8 also being related to immunoglobulin, half of these proteins are related to the immune system (Parra et al., 2016). All of these immunoglobulin proteins were up-regulated in the CALD samples. Also related to the immune system is the CD5 molecule like (CD5L) protein, a secreted glycoprotein that participates in host response to bacterial infection (Sanjurjo et al., 2015) and is also known to regulate lipid biosynthesis (Wang et al., 2015). Beyond immune system proteins, we identified proteins associated with the brain or with involvement in lipid metabolism. Extracellular matrix protein 1 (ECM1) has been associated with lipoid proteinosis in which brain damage develops over time and is associated with the development of cognitive disabilities and epileptic seizures (Zhang et al., 2014). The role of apolipoprotein L1 (APOL1) is not yet clear but it has been associated with the lipid biology in the podocyte





(Fornoni et al., 2014). Copy number variants of the multiple inositol-polyphosphate phosphatase 1 (*MINPP1*) gene have been associated with varying levels of inositol hexaphosphate (IP6)

(Waugh, 2016) and IP6 has been reported to suppress lipid peroxidation (Foster et al., 2017). Apolipoprotein 3 (APOC3) is a key player in triglyceride-rich lipoprotein metabolism

(Ramms and Gordts, 2018) and regulated by the peroxisome proliferator-activated receptor- $\alpha$  (Liu et al., 2015). Lastly, profilin 1 (PFN1) has recently been reported in a CALD study which looked at markers of autoreactivity, identifying anti-PFN1 antibodies present in a large proportion of CALD patients (Orchard et al., 2019b). Together, these protein signals could have significance with respect to the pathophysiology of cerebral demyelination, by highlighting differences around proteins involved in lipid metabolism as well as immune response.

### Estimating Variance of Family Effect

The univariate modeling of CALD vs. non-CALD for each of the individual omics platforms was unsuccessful in identifying significant hits after multiple testing correction. While traditional multiple testing correction methods may be too strict for the omics technologies, we still cannot rule out the possibility that our top hits arise by chance due to variability. Furthermore, our top hits per platform still exhibited a high amount of variance between families, and a lack of consistent signal in molecular features across the entire cohort (Figures 4B,D,F). Within our model we included the effect of the family on the level of the measured signal, and thus we are able to capture the contribution of family structure to a feature's abundance (Eq. 1 and Eq. 2). To illustrate the contribution of these effects, we partitioned the variance contribution within our linear models (Methods). The phenotype effect, total family effect, and residual variance were extracted from our model for each of the features within the RNA-seq, proteomics, and lipidomics platforms (Supplementary Figure S9). As DNAm varies with age we additionally extracted the variance contributed from the age or phenotype-by-age effects. Clearly, the contribution of variance from the phenotype is small in the majority of features across all omics datasets, and a large residual variance indicates a high level of noise present in these high dimensional assays (Supplementary Figure S9). We further demonstrated the heterogeneity in the data by subsetting the families and then repeating comparisons between CALD and non-CALD phenotypes. By leaving out one or two families, the  $\beta$  in equations 1–2 are re-evaluated for the RNA, protein, and lipid datasets. The number of candidates increased with removal of each family, which could be interpreted as potential modifier signatures present in a subset of families, but absent from others (Supplementary Figure S10).

### Integrating Multi-Omic Datasets

As it was our intention to identify molecular marker features underlying cerebral demyelination, we investigated the omics datasets independently to identify a consistent signal. However, owing to a large amount of inter-family variance, we are limited in our ability to identify a statistically significant feature which separates the two phenotypes. As the multi-omic assays should be complementary to each other, we searched for genes which showed differences between the groups in multiple assays. We searched the phenotype comparison between all families, as well as the results from the leave-one-out analysis, wherein we withheld a family and repeated the modeling between the two phenotype groups (Methods). Intersections showed overlapping evidence at the DNA methylation and RNA levels, as well

as overlap between RNA and protein levels, for eight genes. Focusing only on the intersection of all families, only *PTPRN2* has differential signal from both DNA methylation and RNA levels (Figures 4B,D). Additional genes were identified in the leave-one-out subsets (Table 2).

In the multi-omic data we observed that several of the molecular features have trends of differential abundance/expression in a subset of the families. To illustrate this, and attempt to identify clusters within the data, we gathered per-family log<sub>2</sub>-fold-change of CALD over non-CALD for the top hits from the lipid, protein, and RNA datasets. We took this approach because it removes the differences in absolute levels of expression between families. Noticeably, the fold-change values are not consistent for each family, as evidenced by a lack of consistent coloring for each of the features (rows) within the heatmap (Supplementary Figure S11A). Family 2 and family 6 were more similar in their CALD/non-CALD ratios for these features. This is further supported by a principal component analysis, wherein family 2 and family 6 are separated from the other four families on the first principal component (Supplementary Figure S11C). However, this trend does not hold when the set of features is increased to all hits with  $p$ -value < 0.05 across the three platforms, as family 1 and 5 cluster together with the other four families as an outer group (Supplementary Figure S11B). Thus, clustering these families based on top differential features does not reveal confident sub-groupings within the small cohort.

### Specific Modifier Hypothesis Testing

Finding molecular markers which delineate cerebral demyelination in patients with ALD is an ongoing research problem. Additionally, understanding the pathophysiology of cerebral demyelination and potential disruption of the blood brain barrier has implications for diseases beyond ALD. Different hypotheses have been suggested, including involvement of the immune system in autoreactivity or as a response to severe viral

**TABLE 2** | Intersections of significant hits from multiple platforms.

Comparison	DNAm and RNA	DNAm and Protein	RNA and Protein	DNAm and RNA and Protein
all_families	PTPRN2(↑ - ↓)	-	-	-
wo_fam_1	-	-	-	-
wo_fam_2	-	-	-	-
wo_fam_3	HLA-DQB1(↓ - ↓), IL5RA(↓ - ↑), KIF19 (↑ - ↓)	-	-	-
wo_fam_4	-	-	-	-
wo_fam_5	-	-	ICAM1(↑ - ↓), APOL1(↑ - ↑), CD14(↑ - ↑)	-
wo_fam_6	-	-	JCHAIN(↑ - ↑)	-

For each comparison including all families, and each possible 5 × 5 comparison between CALD and non-CALD, the significant hits ( $p$ -value < 0.05 before multiple testing correction) from DNA methylation (DNAm), RNA-sequencing (RNA), and protein LCMS (Protein) were intersected. (↑ means up-regulated/higher for CALD, ↓ means lower in CALD). Intersections with no hits are marked by '-.'

infections. Using the multi-omics dataset, which gives us insight into the complexities of the underlying complex biological system, we tested recently proposed modifiers of cerebral demyelination to see if there is evidence of their discriminatory power within the blood samples profiled in our dataset.

It has recently been demonstrated that autoreactivity to PFN1 occurs in patients affected by CALD, and may be a discriminating marker of cerebral demyelination (Orchard et al., 2019b). We investigated differences in PFN1 methylation, RNA, and protein levels between CALD and non-CALD patients to see if this observation is confirmed in our dataset. At the methylation and RNA level, we did not see a consistent signal differentiating the CALD and non-CALD groups, but at the protein level we observe an increased amount of PFN1 in the CALD group for four out of six families (Figures 4F, 5A,B). This is consistent with the observation from the previous study that not all patients exhibit PFN1 autoreactivity, and the increased protein levels could precede or act as biomarkers of the autoimmune response within the subset of patients who exhibit this trend. While autoreactivity and increased protein abundance are not equivalent, the authors of the previous study showed significant increase in PFN1 protein abundance in the cerebrospinal fluid.

Another study focused on DNA methylation (DNAm) as a marker of CALD, and investigated the intact white matter of brains from patients affected by ALD with and without the cerebral demyelination phenotype (Schlüter et al., 2018). Whether or not the signals they identify confirm within the blood within a separate cohort is important if these proposed marker genes are to be used within newborn screening. Within their analysis they identified differential methylation signals at several genes, two of which are lipin 1 (*LPIN1*) and unc-45 myosin chaperone A (*UNC45A*). Within this cohort, we see no differential methylation signal in the blood for *LPIN1*, and a slight hypermethylation (although not significant) in *UNC45A* (Figures 5C,E). Investigating the RNA shows that while both these genes are highly expressed, there are no consistent differences between the two phenotype groups (Figures 5D,F).

Lastly, it is possible that a viral infection causing an immune response is the phenotypic trigger for progression to CALD, as this is suggested to be a candidate environmental modifier from other cerebral demyelination diseases including multiple sclerosis (Libbey et al., 2014). As is the case in several cancers, RNA-seq can capture actively expressing viral RNA within a sample. To test the hypothesis of whether or not we could observe different expressing viruses within the RNA-seq of these patients, we used the tool Centrifuge to identify traces of viral (or bacterial) sequences (Supplementary Table S10) (Kim et al., 2016). Aside from identifying human, synthetic construct, and endogenous retrovirus, no significant viral or bacterial sequences were identified.

## DISCUSSION

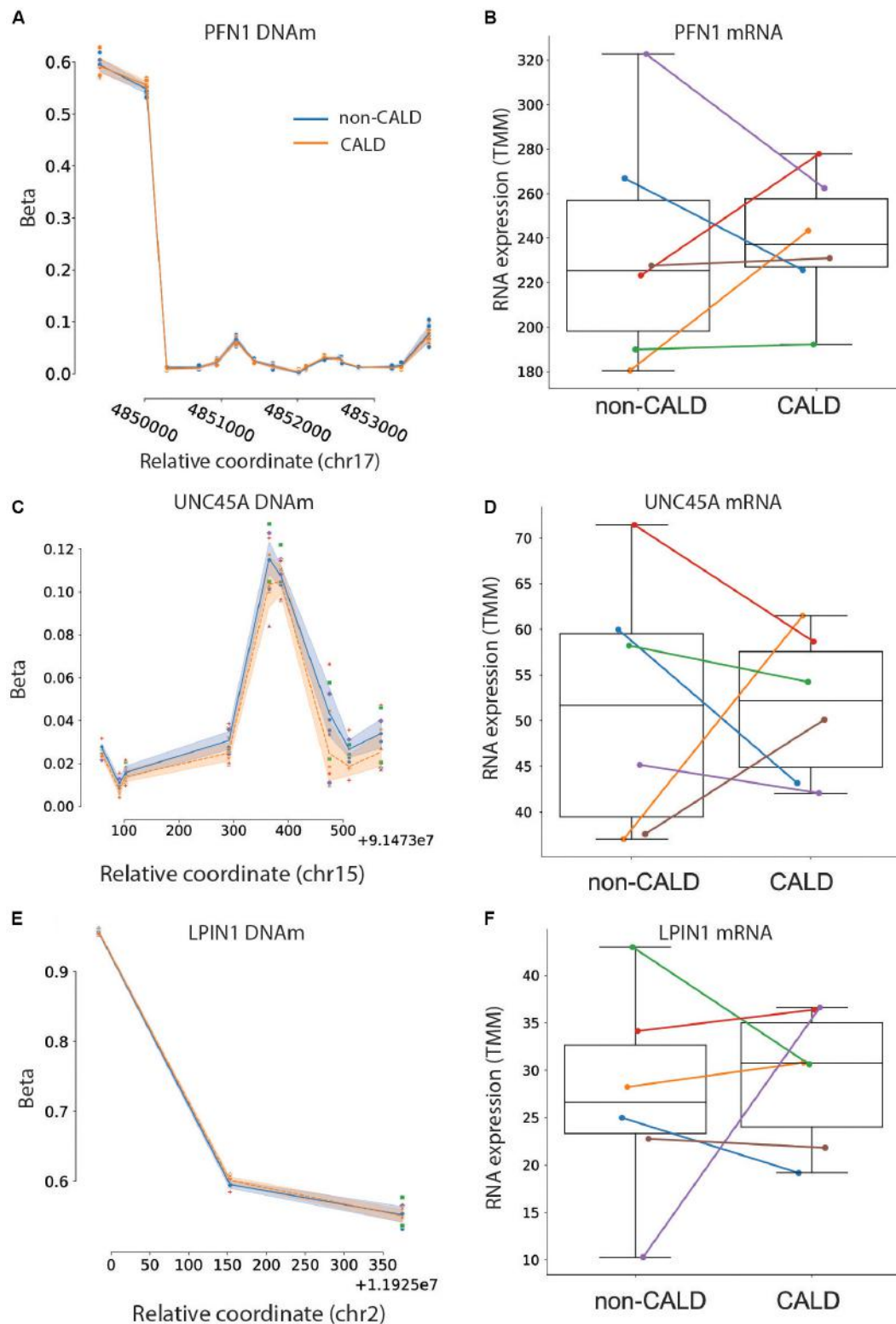
In this study we took a systems biology approach to identify personal molecular characteristics, either genetic or molecular

markers, which may prognosticate the onset of cerebral demyelination in patients affected by ALD. Identifying a single modifier consistent across all individuals has importance because of its potential utility as prognosticator or biomarker heralding the transition to cerebral demyelination, and this carries tremendous treatment implications.

Our cohort was comprised of carefully phenotyped brothers affected by ALD who were discordant for the severe cerebral demyelination phenotype. We collected blood and performed high throughput experiments to profile the DNA, methylated DNA, RNA, lipids, and proteins. In summary, we did not find a strong, convincing, univariate marker which can differentiate all of the CALD and non-cerebral patients in this small cohort. There are several explanations for this negative result: the small cohort with only six discordant sibling pairs of different ethnic background, the possibility that one or more non-CALD patients may still develop cerebral demyelination, high inter-individual variability, and finally the possibility of multiple modifiers and/or an exogenous or non-genetic modifier such as infection or physical trauma. In spite of these limitations, we still emerged with interesting results from each of the omics platforms from this pilot study including discordant genotypes separating all CALD and non-CALD patients, confirmations of recently proposed CALD modifiers, and a suspected involvement of differential activity within the immune system in patients with cerebral demyelination.

In our genetic approach, we identified two discordant genotypes shared between all six brother-pairs: an intronic SNV in *DEU1* and an SNV downstream of *WIBG*. Although *in silico* analysis of the variants and the function of the associated genes did not link these alleles to the cerebral demyelination phenotype, it is of interest to see if they replicate in a larger cohort. Examining variants shared by all CALD patients led to the identification of a missense polymorphism in *TPCN2*, a gene which localizes to lysosomal membranes. This exists in a segregating haplotype block, and is absent from all non-CALD patients except for family 4 – the youngest patient with the highest chance to develop cerebral demyelination – where the haploblock is homozygous. How this variant segregates in a larger patient cohort could be of interest. None of the previously proposed modifier alleles, emerging from GWAS or target-gene studies, confirmed within our cohort.

Although our analysis was burdened by high inter-individual variability, we were able to identify univariate molecular markers with increased confidence due to replication—by multiple omics levels and/or by confirming previously proposed modifier markers. A recent study Schlüter et al. (2018) showed CALD patients with DNA hypermethylation within *PTPRN2*, which we confirm in our study and support with decreased mRNA expression in CALD patients (both platforms reaching  $p$ -value < 0.05 before multiple testing correction). The same study showed hypermethylation of *LPIN1* and *UNC45A*, the latter of which we confirm (although not statistically significant) as slightly hypermethylated in CALD samples. Of note, that study used brain tissue to derive their signal whereas we use blood samples. Another study utilized CSF and blood plasma, including longitudinal data from ALD patients pre- and post



**FIGURE 5 |** Testing previously suggested markers. DNA methylation and mRNA abundance for PFN1 (A,B), UNC45A (C,D), and LPIN1 (E,F). DNA methylation is shown for all CpGs associated to the listed genes, with the non-CALD mean methylation shown as a blue line with standard error shading, and the orange dashed line showing the mean methylation of CALD. Individual points are shown and colored by family. RNA expression is shown as a boxplot for non-CALD and CALD phenotype groups, with individual families labeled with family 1 as blue, family 2 as orange, family 3 as green, family 4 as red, family 5 as purple, and family 6 as brown.



cerebral demyelination, to identify autoreactivity to Profilin 1 (PFN1) within CALD patients (Orchard et al., 2019b). They observed auto-antigens to PFN1 in the blood, and increased PFN1 levels in CSF, in ~50% of CALD patients. In our cohort, four out of six patients exhibit increased PFN1 protein levels, in-line with the observation that PFN1 phenotype is not ubiquitous across all CALD patients. While increased abundance of PFN1 in the plasma was not investigated in the previous study, it is encouraging to see the higher levels of PFN1 in the plasma of the CALD patients within this cohort. We further contribute to this observation by showing no differences at the DNAm or mRNA levels, pointing toward a separate mechanism of upregulation/overabundance of PFN1. The role of PFN1 overabundance, and its potential link to autoreactivity, is a potential avenue of further research should these findings confirm in a larger cohort.

As ALD is a peroxisomal disorder, the lipidomic analysis presented here is of interest. The lipid profiling data confirmed previous observations regarding VLCFA abundance differences in ALD samples when compared to controls. Specifically, the phosphatidylcholines (PC) species containing very long-chain fatty acids are more abundant in the ALD group compared to the control group. Furthermore, the suitability of LPC(C26:0) to function as a marker for ALD in newborn screening was confirmed. Differences in lipid abundance between CALD and non-CALD groups did not reach significance after multiple testing correction, likely due to a lack of consistent lipid differences between all brother pairs. Nevertheless, the chain lengths and position of the double bonds of the differential lipids between CALD and non-CALD could provide insight into the pathophysiology of CALD as CALD patients had lower levels of sphingomyelin and its precursor ceramide, in line with disease progression (**Supplementary Figure S1**). The specificity of the lipids into play is further illustrated by the fact that for some ceramides (d43:3) the control group has higher abundances than ALD samples while for other ceramides (d46:4) the abundances are much lower. This specificity of individual sphingolipids in CALD could be of interest as they have previously been demonstrated to be relevant in other peroxisomal disorders (Herzog et al., 2018a,b).

Beyond identifying phenotype-stratifying molecular features, we investigated the top hits at the gene-level from each omics platform for any relation to the pathophysiology of CALD. Literature searches highlighted genes involved in lipid metabolism, the nervous system, and the immune system. Gene Ontology and KEGG pathways further supported these observations. While our analysis highlights the potential for these systems to be affected, we are hesitant based on the available data to draw conclusions about the impact of cerebral demyelination with respect to the inflammatory system. The inflammatory response and its link to cerebral demyelination is the subject of ongoing research, including investigation of the initiation and maintenance of inflammation and its contribution to clinical features (van der Voorn et al., 2011; Musolino et al., 2012). Larger datasets, and datasets of relevant affected tissues, are needed to draw conclusions from differentially abundant molecular features.

Throughout this work we have identified certain limitations of our approach which should be considered in future work focused on modifiers of rare disease, especially for other inborn errors of metabolism (e.g., Gaucher disease). First, we suffered from having a small number of samples and a high number of observed features. For future univariate marker investigations we recommend focusing only on protein or mRNA and increasing the number of samples. Second, our genetic analysis was limited by the possibility of future transition to the CALD state for any of our non-CALD patients, especially those patients who have not reached maturity. Recent epidemiological analysis shows that cerebral demyelination can occur throughout the lifetime of an ALD patient (Huffnagel et al., 2019b), so genetic studies should focus on older (60–70 years old) patients who have not developed the cerebral demyelination phenotype. While discordant brother pairs reaching old age are challenging to find, a collection of genotyped non-CALD patients older than 60–70 years of age could serve as a good control. Third, we are limited in capturing relevant biological insights because we are profiling blood and not CSF/brain tissue. We recognize that lymphocytes and plasma are not ideal to study the pathophysiology of cerebral demyelination, as that would require samples from the white matter of the brain or an *in vitro* model system (e.g., differentiated iPSCs). We chose to use blood-based samples in this project as our goal was to find a modifier signal or molecular marker from an accessible source within affected patients. There are future opportunities for multi-omic profiling of these more relevant samples to identify the molecular underpinnings of cerebral demyelination. Lastly, while we profile DNA methylation, we don't capture other components of the environment which could have an impact including microbiome and pathogen exposure history.

With newborn screening now a reality for ALD, prognostication and timing of therapy becomes more relevant than ever before; thus modifier studies to decipher a protector or marker for cerebral demyelination will continue (Moser and Fatemi, 2018). We believe that this dataset can continue to be mined and used for testing the replication of proposed phenotypic markers. Further, the data within this study could be used as part of a larger dataset examining multivariate signals differentiating the two classes. Whether it is a collection of genetic markers or a pattern of multiple molecular features, it is clear that there is a need for a larger sample size. Additional samples will enable statistical bootstrapping, and allow for additional insights to be collected from the profiled samples. As such, we make the measurements within this study available for future use to the community (see section "Data Availability"), with the hopes that the data can serve as a secondary confirmation of new modifier hypotheses, or as part of a larger dataset for investigating the complex nature of cerebral demyelination.

## DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in the following repositories: Zenodo, 10.5281/zenodo.382068; Github (code): [https://github.com/Phillip-a-richmond/ALD\\_Modifier\\_Project](https://github.com/Phillip-a-richmond/ALD_Modifier_Project).

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the PEROX Biobank according to a protocol (METC2015\_066) approved by the biobank review board of the Amsterdam UMC (BioBank Toetsingscommissie AMC). All patients provided written informed consent for storage and use of materials for medical research. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

## AUTHOR CONTRIBUTIONS

Project was conceived by SK, ME, and CK. Patients were recruited by ME. Manuscript written by PR, FK, WW, CK, SK, and ME. Data analysis for this work includes: WGS data analysis by PR. RNA-seq analysis by FK. Lipidomics analysis by FV. Proteomics analysis by AU and PL. DNA methylation analysis by DL and MK. Statistical analysis by EG, SM, PM, and AK. All authors contributed to the article and approved the submitted version.

## REFERENCES

- Argmann, C. A., Houten, S. M., Zhu, J., and Schadt, E. E. (2016). A next generation multiscale view of inborn errors of metabolism. *Cell Metab.* 23, 13–26. doi: 10.1016/j.cmet.2015.11.012
- Barendsen, R. W., Dijkstra, I. M. E., Visser, W. F., Alders, M., Blik, J., and Boelen, A. (2020). Adrenoleukodystrophy newborn screening in the Netherlands (SCAN Study): the X-Factor. *Front. Cell Dev. Biol.* 8:499. doi: 10.3389/fcell.2020.00499
- Bunker, R. D., Bulloch, E. M. M., Dickson, J. M. J., Loomes, K. M., and Baker, E. N. (2013). Structure and function of human xylulokinase, an enzyme with important roles in carbohydrate metabolism. *J. Biol. Chem.* 288, 1643–1652. doi: 10.1074/jbc.M112.427997
- de Beer, M., Engelen, M., and van Geel, B. M. (2014). Frequent occurrence of cerebral demyelination in adrenomyeloneuropathy. *Neurology* 83, 2227–2231. doi: 10.1212/WNL.0000000000001074
- Engelen, M., Barbier, M. I., Dijkstra, M., Schur, R., de Bie, R. M., Verhamme, C., et al. (2014). X-linked adrenoleukodystrophy in women: a cross-sectional cohort study. *Brain* 137(Pt 3), 693–706. doi: 10.1093/brain/awt361
- Engelen, M., Kemp, S., de Visser, M., van Geel, B. M., Wanders, R. J., Aubourg, P., et al. (2012). X-linked adrenoleukodystrophy (X-ALD): clinical presentation and guidelines for diagnosis, follow-up and management. *Orphanet J. Rare Dis.* 7:51. doi: 10.1186/1750-1172-7-51
- Fornoni, A., Merscher, S., and Kopp, J. B. (2014). Lipid biology of the podocyte—new perspectives offer new opportunities. *Nat. Rev. Nephrol.* 10, 379–388. doi: 10.1038/nrneph.2014.87
- Foster, S. R., Dilworth, L. L., Omoruyi, F. O., Thompson, R., and Alexander-Lindo, R. L. (2017). Pancreatic and renal function in streptozotocin-induced type 2 diabetic rats administered combined inositol hexakisphosphate and inositol supplement. *Biomed. Pharmacother.* 96, 72–77. doi: 10.1016/j.biopha.2017.09.126
- Génin, E., Feingold, J., and Clerget-Darpoux, F. (2008). Identifying modifier genes of monogenic disease: strategies and difficulties. *Hum. Genet.* 124, 357–368. doi: 10.1007/s00439-008-0560-2
- Harrison, X. A., Donaldson, L., Correa-Cano, M. E., Evans, J., Fisher, D. N., Goodwin, C. E. D., et al. (2018). A brief introduction to mixed effects modelling and multi-model inference in ecology. *PeerJ* 6:e4794.
- Herzog, K., Pras-Raves, M. L., Ferdinandusse, S., Vervaart, M. A. T., Luyf, A. C. M., van Kampen, A. H. C., et al. (2018a). Functional characterisation of peroxisomal

## FUNDING

PR was supported by BC Children's Hospital Research Institute Graduate Studentship. ME was supported by an NWO/ZonMW Vidi grant (016.196.310). CK was supported by a Stichting Metakids grant. This research project was funded by Stichting Steun Emma Kinderziekenhuis (Amsterdam; project number WAR 2016-014).

## ACKNOWLEDGMENTS

We gratefully acknowledge the patients and families for their participation in this study, and the clinicians and colleagues for their expert management.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fcell.2020.00520/full#supplementary-material>

- beta-oxidation disorders in fibroblasts using lipidomics. *J. Inherit. Metab. Dis.* 41, 479–487. doi: 10.1007/s10545-017-0076-9
- Herzog, K., Pras-Raves, M. L., Ferdinandusse, S., Vervaart, M. A. T., Luyf, A. C. M., van Kampen, A. H. C., et al. (2018b). Plasma lipidomics as a diagnostic tool for peroxisomal disorders. *J. Inherit. Metab. Dis.* 41, 489–498. doi: 10.1007/s10545-017-0114-7
- Hoffman, G. E., and Schadt, E. E. (2016). variancePartition: interpreting drivers of variation in complex gene expression studies. *BMC Bioinformatics* 17:483. doi: 10.1186/s12859-016-1323-z
- Hoffmann, H.-H., Schneider, W. M., and Rice, C. M. (2015). Interferons and viruses: an evolutionary arms race of molecular interactions. *Trends Immunol.* 36, 124–138. doi: 10.1016/j.it.2015.01.004
- Hosseinibarkooie, S., Peters, M., Torres-Benito, L., Rastetter, R. H., Hupperich, K., Hoffmann, A., et al. (2016). The power of human protective modifiers: PLS3 and CORO1C unravel impaired endocytosis in spinal muscular atrophy and rescue SMA phenotype. *Am. J. Hum. Genet.* 99, 647–665. doi: 10.1016/j.ajhg.2016.07.014
- Huffnagel, I. C., Dijkgraaf, M. G. W., Janssens, G. E., van Weeghel, M., van Geel, B. M., Poll-The, B. T., et al. (2019a). Disease progression in women with X-linked adrenoleukodystrophy is slow. *Orphanet J. Rare Dis.* 14:30.
- Huffnagel, I. C., Laheji, F. K., Aziz-Bose, R., Tritos, N. A., Marino, R., Linthorst, G. E., et al. (2019b). The natural history of adrenal insufficiency in X-linked adrenoleukodystrophy: an international collaboration. *J. Clin. Endocrinol. Metab.* 104, 118–126. doi: 10.1210/je.2018-01307
- Huffnagel, I. C., van Ballegoij, W. J. C., van Geel, B. M., Vos, J. M. B. W., Kemp, S., and Engelen, M. (2019c). Progression of myelopathy in males with adrenoleukodystrophy: towards clinical trial readiness. *Brain* 142, 334–343. doi: 10.1093/brain/awy299
- Kemp, S. I., Huffnagel, C., Linthorst, G. E., Wanders, R. J., and Engelen, M. (2016). Adrenoleukodystrophy – neuroendocrine pathogenesis and redefinition of natural history. *Nat. Rev. Endocrinol.* 12, 606–615. doi: 10.1038/nrendo.2016.90
- Kemp, S., Berger, J., and Aubourg, P. (2012). X-linked adrenoleukodystrophy: clinical, metabolic, genetic and pathophysiological aspects. *Biochim. Biophys. Acta* 1822, 1465–1474. doi: 10.1016/j.bbdis.2012.03.012
- Khan, A., and Mathelier, A. (2017). Intervene: a tool for intersection and visualization of multiple gene or genomic region sets. *BMC Bioinformatics* 18:287. doi: 10.1186/s12859-017-1708-7

- Kim, D., Song, L., Breitwieser, F. P., and Salzberg, S. L. (2016). Centrifuge: rapid and sensitive classification of metagenomic sequences. *Genome Res.* 26, 1721–1729. doi: 10.1101/gr.210641.116
- Korenke, G. C., Fuchs, S., Krasemann, E., Doerr, H. G., Wilichowski, E., Hunneman, D. H., et al. (1996). Cerebral adrenoleukodystrophy (ALD) in only one of monozygotic twins with an identical ALD genotype. *Ann. Neurol.* 40, 254–257. doi: 10.1002/ana.410400221
- Lee, S. E., Simons, S. B., Heldt, S. A., Zhao, M., Schroeder, J. P., Vellano, C. P., et al. (2010). RGS14 is a natural suppressor of both synaptic plasticity in CA2 neurons and hippocampal-based learning and memory. *Proc. Natl. Acad. Sci. U.S.A.* 107, 16994–16998. doi: 10.1073/pnas.1005362107
- Libbey, J. E., Lane, T. E., and Fujinami, R. S. (2014). Axonal pathology and demyelination in viral models of multiple sclerosis. *Discov. Med.* 18, 79–89.
- Liu, C., Guo, Q., Lu, M., and Li, Y. (2015). An experimental study on amelioration of dyslipidemia-induced atherosclerosis by clematichinenoside through regulating peroxisome proliferator-activated receptor- $\alpha$  mediated apolipoprotein A-I, A-II and C-III. *Eur. J. Pharmacol.* 761, 362–374. doi: 10.1016/j.ejphar.2015.04.015
- McEwen, L. M., Jones, M. J., Lin, D. T. S., Edgar, R. D., Husquin, L. T., MacIsaac, J. L., et al. (2018). Systematic evaluation of DNA methylation age estimation with common preprocessing methods and the Infinium methylationEPIC BeadChip array. *Clin. Epigenetics* 10:123. doi: 10.1186/s13148-018-0556-2
- Moser, A., Jones, R., Hubbard, W., Tortorelli, S., Orsini, J., Caggana, M., et al. (2016). Newborn Screening for X-Linked Adrenoleukodystrophy. *Int. J. Neonatal Screen.* 2:15. doi: 10.3390/ijns2040015
- Moser, A. B., and Fatemi, A. (2018). Newborn screening and emerging therapies for X-linked adrenoleukodystrophy. *JAMA Neurol.* 75, 1175–1176.
- Musolino, P. L., Rapalino, O., Caruso, P., Caviness, V. S., and Eichler, F. S. (2012). Hypoperfusion predicts lesion progression in cerebral X-linked adrenoleukodystrophy. *Brain* 135(Pt 9), 2676–2683. doi: 10.1093/brain/aws206
- Ofman, R. I., Dijkstra, M. E., van Roermund, C. W. T., Burger, N., Turkenburg, M., van Cruchten, A., et al. (2010). The role of ELOVL1 in very long-chain fatty acid homeostasis and X-linked adrenoleukodystrophy. *EMBO Mol. Med.* 2, 90–97. doi: 10.1002/emmm.201000061
- Orchard, P. J., Markowski, T. W., Higgins, L., Raymond, G. V., Nascene, D. R., Miller, W. P., et al. (2019a). Association between APOE4 and biomarkers in cerebral adrenoleukodystrophy. *Sci. Rep.* 9:7858.
- Orchard, P. J., Nascene, D. R., Gupta, A., Taisto, M. E., Higgins, L., Markowski, T. W., et al. (2019b). Cerebral adrenoleukodystrophy is associated with loss of tolerance to proflin. *Eur. J. Immunol.* 49, 947–953. doi: 10.1002/eji.201848043
- Paila, U., Chapman, B. A., Kirchner, R., and Quinlan, A. R. (2013). GEMINI: integrative exploration of genetic variation and genome annotations. *PLoS Comput. Biol.* 9:e1003153. doi: 10.1371/journal.pcbi.1003153
- Parra, D., Korytář, T., Takizawa, F., and Sunyer, J. O. (2016). B cells and their role in the teletost gut. *Dev. Comp. Immunol.* 64, 150–166. doi: 10.1016/j.dci.2016.03.013
- Pierpont, E. I., Eisengart, J. B., Shanley, R., Nascene, D., Raymond, G. V., Shapiro, E. G., et al. (2017). Neurocognitive trajectory of boys who received a hematopoietic stem cell transplant at an early stage of childhood cerebral adrenoleukodystrophy. *JAMA Neurol.* 74, 710–717.
- Poplin, R., Chang, P.-C., Alexander, D., Schwartz, S., Colthurst, T., Ku, A., et al. (2018). A universal SNP and small-indel variant caller using deep neural networks. *Nat. Biotechnol.* 36, 983–987. doi: 10.1038/nbt.4235
- Ramms, B., and Gordts, P. L. S. M. (2018). Apolipoprotein C-III in triglyceride-rich lipoprotein metabolism. *Curr. Opin. Lipidol.* 29, 171–179. doi: 10.1097/mol.0000000000000502
- Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., et al. (2015). Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43:e47. doi: 10.1093/nar/gkv007
- Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140. doi: 10.1093/bioinformatics/btp616
- Sanjurjo, L., Amézaga, N., Aran, G., Naranjo-Gómez, M., Arias, L., Armengol, C., et al. (2015). The human CD5L/AIM-CD36 axis: a novel autophagy inducer in macrophages that modulates inflammatory responses. *Autophagy* 11, 487–502. doi: 10.1080/15548627.2015.1017183
- Schlüter, A., Sandoval, J., Fourcade, S., Iaz-Lagares, A. D., Ruiz, M., Casaccia, P., et al. (2018). Epigenomic signature of adrenoleukodystrophy predicts compromised oligodendrocyte differentiation. *Brain Pathol.* 28, 902–919. doi: 10.1111/bpa.12595
- Sengelaub, C. A., Navrazhina, K., Ross, J. B., Halberg, N., and Tavazoie, S. F. (2016). PTPRN 2 and PLC  $\beta$ 1 promote metastatic breast cancer cell migration through PI (4,5)P 2 -dependent actin remodeling. *EMBO J.* 35, 62–76. doi: 10.15252/emboj.201591973
- Sittaramane, V., and Chandrasekhar, A. (2008). Expression of unconventional myosin genes during neuronal development in zebrafish. *Gene Expr. Patterns* 8, 161–170. doi: 10.1016/j.gexp.2007.10.010
- Turer, E., McAlpine, W., Wang, K.-W., Lu, T., Li, X., Tang, M., et al. (2017). Creatine maintains intestinal homeostasis and protects against colitis. *Proc. Natl. Acad. Sci. U.S.A.* 114, E1273–E1281.
- van der Voorn, J. P., Pouwels, P. J., Powers, J. M., Kamphorst, W., Martin, J. J., Troost, D., et al. (2011). Correlating quantitative MR imaging with histopathology in X-linked adrenoleukodystrophy. *AJNR Am. J. Neuroradiol.* 32, 481–489. doi: 10.3174/ajnr.a2327
- van Engen, C. E., Ofman, R. I., Dijkstra, M. E., van Goethem, T. J., Verheij, E., Varin, J., et al. (2016). CYP4F2 affects phenotypic outcome in adrenoleukodystrophy by modulating the clearance of very long-chain fatty acids. *Biochim. Biophys. Acta* 1862, 1861–1870. doi: 10.1016/j.bbdis.2016.07.006
- Wang, C., Yosef, N., Gaublot, J., Wu, C., Lee, Y., Clish, C. B., et al. (2015). CD5L/AIM Regulates Lipid Biosynthesis and Restrains Th17 Cell Pathogenicity. *Cell* 163, 1413–1427. doi: 10.1016/j.cell.2015.10.068
- Waugh, M. G. (2016). Chromosomal instability and phosphoinositide pathway gene signatures in glioblastoma multiforme. *Mol. Neurobiol.* 53, 621–630. doi: 10.1007/s12035-014-9034-9
- Wiesinger, C., Eichler, F. S., and Berger, J. (2015). The genetic landscape of X-linked adrenoleukodystrophy: inheritance, mutations, modifier genes, and diagnosis. *Appl. Clin. Genet.* 8, 109–121.
- Willrich, V. M. A., Kaleta, E. J., Bryant, S. C., Spears, G. M., Train, L. J., Peterson, S. E., et al. (2018). Genetic variation in statin intolerance and a possible protective role for UGT1A1. *Pharmacogenomics* 19, 83–94. doi: 10.2217/pgs-2017-0146
- Zhang, R., Liu, Y., Xue, Y., Wang, Y., Wang, X., Shi, S., et al. (2014). Treatment of lipid proteinosis due to the p.C220G mutation in ECM1, a major allele in Chinese patients. *J. Transl. Med.* 12:85. doi: 10.1186/1479-5876-12-85

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The handling editor declared a past co-authorship with one of the authors SK.

Copyright © 2020 Richmond, van der Kloet, Vaz, Lin, Uozie, Graham, Kobor, Mostafavi, Moerland, Lange, van Kampen, Wasserman, Engelen, Kemp and van Karnebeek. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.