**INTRODUCTION** Adrenoleukodystrophy (ALD) is a rare peroxisomal X-linked degenerative disease (MIM 300100), caused by deficiency of the ABC transporter encoded by *ABCD1*[1] (Fig 1). More than 700 different disease-causing *ABCD1* mutations have been reported, all leading to very long chain fatty acid accumulation, adrenal insufficiency and myelin destruction. The overall incidence is 1:17,000. In males, ALD often manifests with adrenocortical insufficiency in childhood (80% before 18 years). During adulthood virtually all male (and eventually) female patients develop a progressive myelopathy (*Adrenomyeloneuropathy).* Additionally, male patients can develop cerebral demyelination (*Cerebral ALD)* in childhood, but also in adulthood. It is estimated that eventually more than 60% of male patients develop cerebral ALD. Untreated cerebral ALD is progressive and causes vegetative state and death 2-3 years after onset. *There is no correlation between genotype and symptoms.* Even siblings with the same mutation can have vastly different outcomes (death at 4yrs *vs* living into adulthood), so something more is at the root of ALD severity. Cerebral ALD can be treated by hematopoietic stem cell transplant (HSCT), but only in the pre- or early symptomatic state. Newborn screening (NBS) for ALD has begun, enabling ALD diagnosis at birth, but unless we can predict the disease course, clinical decision-making will remain symptom-dependent. We hypothesize that we can improve outcome-prediction for ALD by identifying modifiers of disease severity in phenotypically discordant siblings using a systems biology approach (*see Figure 1*).

**OBJECTIVES:** Using integrated bioinformatics analysis of data from 35 exquisitely phenotyped brothers with discordant ALD symptoms, we will:

1. Use whole genome sequencing (WGS) to identify candidate genetic and viral/pathogen modifiers.
2. Profile DNA methylation in peripheral blood to identify epigenetic modifiers.
3. Use lipidomics to identify/validate metabolic markers.
4. Validate candidate modifiers/ markers in unrelated ALD patients and iPS cells.

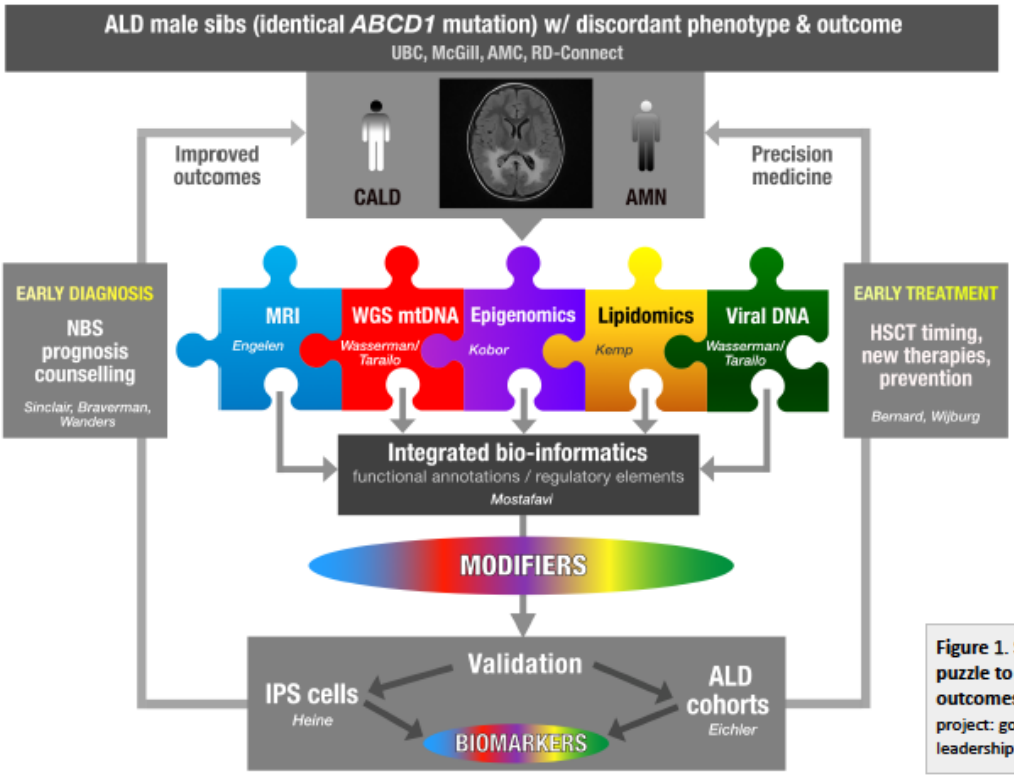


**Figure 1. Solving the ALD puzzle to improve health outcomes.** An overview of project: goals, activities, leadership and deliverables.

This study is the first of its kind for ALD and can serve as a model for other inborn error of metabolisms (IEMs) with similar phenotypic diversity and inherent management challenges.

## BACKGROUND & RATIONALE

Our research focuses on delineating phenotypic variability in ALD, of which the biochemical pathophysiology is depicted in Fig 2.

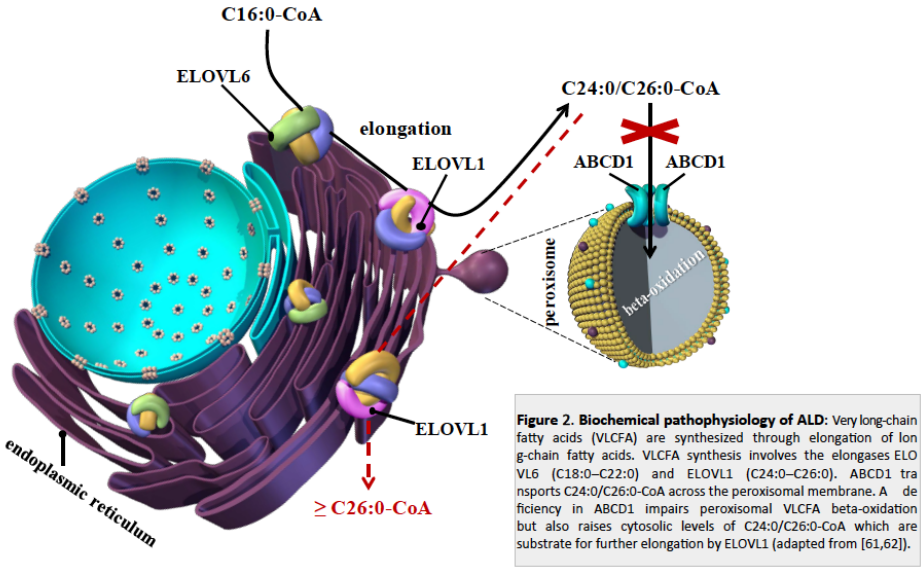


**Figure 2. Biochemical pathophysiology of ALD:** Very long-chain fatty acids (VLCFA) are synthesized through elongation of long-chain fatty acids. VLCFA synthesis involves the elongases ELOVL6 (C18:0–C22:0) and ELOVL1 (C24:0–C26:0). ABCD1 transports C24:0/C26:0-CoA across the peroxisomal membrane. A deficiency in ABCD1 impairs peroxisomal VLCFA beta-oxidation but also raises cytosolic levels of C24:0/C26:0-CoA which are substrate for further elongation by ELOVL1 (adapted from [61,62]).

There are ~250 births affected by this neurodegenerative disease in Canada and the US each year, with two predominant phenotypes:[60]

- Adrenomyeloneuropathy (AMN): the slowly progressive adult onset form, is characterized by axonopathy; all adult males with ALD will develop AMN; most females also develop myelopathy.[2]
- Cerebral ALD (CALD): the rapidly progressive cerebral form of early onset, is characterized by fatal inflammatory demyelination in brain[3]. CALD is seen only in males[4] and, if detected early, can be arrested by HSCT[5]; preventive HSCT is not an option given high morbidity and mortality.

The advent of NBS for ALD will identify boys who are asymptomatic at birth, and ensure they receive long term follow-up and monitoring. To ensure the timely HSCT for males with CALD, they are subjected to rigorous neurological and MRI follow-up that poses considerable physical, emotional and financial burden. More than 700 non-recurrent ALD-causing mutations in *ABCD1* have been described to date (http://www.x-ald.nl)[6]. Phenotypic discordance in individuals with the same *ABCD1* genotype, including siblings and monozygotic twins[7], strongly supports the hypothesis that other modifying factors play a role in the progression of the disease[8]. As yet, however, modifier studies using both candidate gene and genome-wide association study (GWAS) approaches have had little success and resulted in the identification of only a single modifier gene (*CYP4F2*) with limited predictive power;[26] this is likely because of limited genomic search space that was explored; [8] which can potentially be overcome by a systems biology approach.

In multiple sclerosis (MS), another progressive white matter disease –albeit non-monogenic-, external pathogens may cause inflammatory cerebral demyelination[9,10] and evidence supports an association of Epstein-Barr virus with onset and/or relapse in MS. Such infections could promote a pro-inflammatory state in CALD[9]; this possibility will be addressed for the first time as part of our project.

The unresolved phenotypic variability of ALD is a crucial roadblock for patient care. Our goal is to delineate *personal molecular characteristics* that contribute to phenotypic variability (AMN vs CALD)

in male ALD siblings to define biomarkers that predict onset and progression of CALD. This represents a novel application of an approach we have successfully used for discovery of other treatable IEMs in our patients[11–20]. This project is an interdisciplinary and international initiative that combines clinical expertise in deep phenotyping of ALD patients with research expertise in genomics, epigenomics, lipidomics and bioinformatics. We expect to identify epi/genetic modifiers and possible viral triggers of ALD that we may use as biomarkers/predictors of CALD in ALD patients. As such, this work allows for Precision Medicine by providing better pathophysiologic understanding and prediction of disease course for the affected individual, allowing for improved health management decisions about timing and choice of treatments. With recent technical advancements allowing for inclusion of ALD in routine NBS programs, and the associated stakeholder push for this addition, there is immediate need for reliable predictors of disease progression. Furthermore, the discovery of pathways that lead to attenuation of ALD phenotype may reveal new therapeutic targets, exemplified by another rare devastating neurologic disease, Spinal Muscular Atrophy (MIM 253300), for which PLS3 and CORO1C were identified as protective modifiers, unravelling impaired endocytosis as rescue mechanism for its SMA Phenotype.[64] Finally, this project nurtures unique skillsets in this and the next generation of clinical and basic scientists to apply a systems biology approach to advance health care.

## PREVIOUS WORK

This interdisciplinary collaborative project between Canada (University of British Columbia, McGill University), the Netherlands (Academic Medical Centre, Free University Medical Centre in Amsterdam) and the United States (Harvard Medical School) is the first of its kind, focused on unraveling prognosticators and treatment targets of ALD via an integrated *-omics* analysis. Not only does each centre and set of investigators bring unique and complementary expertise to fulfill the goals of this study, also the infrastructure has been to created to ensure a flying start, both in terms of patients and facilities. *NPA* Dr van Karnebeek and co-applicants Drs Wasserman and Tarailo-Graovac recently published their neurometabolic diseases discovery study in the *New England Journal of Medicine*[21]; an advanced bioinformatics pipeline has been created which translates phenomics, genomics and metabolomics data into monogenic diagnoses enabling treatments with impact on patient care and outcomes and it certainly provides the basis for the genomics analysis which is so central in this project. [11–20] Examples include the discovery of carbonic anhydrase VA deficiency treatable with carglumic acid, the lysine restricted diet as novel therapy for pyridoxine dependent epilepsy, and sialic acid supplementation for NANS deficiency[22]. *CoA Dr Mike Kobor* has a solid track record in applying epigenetic analysis to a variety of human monogenic and complex diseases including a recent study on human fetal alcohol spectrum disorder[23]. In 2015, coPA *Dr Engelen* began a prospective cohort study of male ALD patients with detailed clinical data and storage of samples. This "Dutch cohort" currently contains 50 male patients with ALD and is expanding. This is true as well for *CoAs Drs Braverman and Bernard*, experts in leukodystrophies with close connections to worldwide networks such ALD-connect (led by Dr. *Eichler*, www.aldconnect.org); the latter is an international group of patients, advocates and researchers (including our team members) who advance ALD research, awareness and education. CoPA *Kemp* recently reported that the *CYP4F2* polymorphism rs2108622 increases the risk of developing cerebral ALD in Caucasian patients. CYP4F2 is important for omega-oxidation of very long chain fatty acids (VLCFA)[24,25] and the SNP lowers the omega-oxidation capacity[26]. *CoPA Dr Mostafavi* is on the cutting edge of computational integration of big data sets, and has successfully applied network-based approaches for combined genomic analysis neuropsychiatric conditions[27,28]. We will now garner our collective expertise working with patients and families to identify disease modifiers addressing the many questions regarding individual prognosis and timing of life-alerting but risky treatments for a quickly expanding group given the advent of newborn screening for ALD.

## 2.1 APPROACH

Our goal is to improve disease outcome prediction for ALD by identifying modifiers of disease severity, using a systems biology approach. In 35 deeply phenotyped brothers with discordant ALD symptoms, we will pursue the following objectives by performing the listed experiments (see Fig 1)

**OBJECTIVES & EXPERIMENTS**

**1. Enrolment, deep phenotyping (Yrs1-2) & follow-up of discordant ALD siblings (*Yrs 1-4; Engelen, Eichler, Sirrs, van Karnebeek, Bernard*)**

In 2015, coPA *Engelen* began a prospective cohort study of 50 male ALD patients (all alive, age 3 to 69 years) to understand phenotypic variability. All subjects undergo annual clinical evaluation (neurologic and adrenal), quantitative MRI, and storage of samples (plasma, lymphocytes, DNA, RNA, CSF). From this Dutch cohort and ALD Connect which is a global network run out of Harvard Medical School, we have selected 35 deeply phenotyped male sibling pairs who are discordant for age of CALD onset, severity/HSCT indication, or ALD phenotype (CALD *vs* AMN). Neuro-imaging, DNA and lipidomics samples are already available for 20 of the subjects. Once Ethics Board approval is obtained we will proceed with the -omics analyses outlined below with modifier validations.
Outcome: Enrolment, samples, detailed clinical and follow-up date for 35 discordant male ALD sibs

**2- Use WGS to identify candidate genetic and viral pathogen modifiers of disease (Yrs 1-4; *Wasserman, Tarailo-Graovac, Matthews, van Karnebeek*)**

Genomic DNA of discordant sibling pairs will be extracted from peripheral blood samples using standard methods and subjected to WGS (50x coverage using IlluminaHiSeqx Ten sequencer, Macrogen). The paired-end sequence reads will be aligned to the human reference genome, and sequence differences assessed (in nuclear and mitochondrial DNA) between siblings, using our validated bioinformatics approach[15], focusing on intra/interfamilial differences in small- (single nucleotide variants, < 20bp insertions and deletions) and large-scale (copy number and structural variants). Bioinformatics tools are rapidly improving and corresponding pipelines evolving. Our current approach for small-scale variants (1-100bp) includes SAMtools[29], Platypus[30] and GATK HaplotypeCaller[31] variant callers and SNPeff annotations[32]. Custom Perl/Python scripts are used to further prioritize variants according to adequate depth of coverage, zygosity, population frequency and pathogenicity scores as previously described[21]. For mid-range indels and copy number variants, we are using a combination of Lumpy[33] and CNVnator[34]. For complex and large structural variants we are using the MetaSV package, as well as exploring *de novo* assembly followed by variant detection using the Abyss and PAVfinder packages[35,36] Beyond the nuclear genome, we use MToolBox[37] for homo/heteroplasmy annotation and prioritization of mitochondrial variants from WGS data.

- We are interested in variants that are unique or of unique genotype (homozygous vs heterozygous) for each affected sibling. In particular, we are looking for: protein-coding variants [as modifiers could be either rare (minor allele frequency<1%) or polymorphic (>1%)]
- non-coding DNA, including regulatory elements[38]
- variants in genes implicated in pathways suggested by lipidomics profiles (**Obj3**)

Candidate variants will be Sanger-confirmed and compared across all 35 families to detect modifying mechanisms. Recently, mining unmapped sequence reads from human WGS analyses has been used to identify viral pathogens, including Epstein-Barr virus[39]. Accordingly, we propose to test the unmapped ALD sequence data against non-human sequence databases to identify viral sequences as an initial investigation of the possible role for viral pathogens in the promotion of cerebral inflammation in CALD. Outcomes: From this objective, we expect to identify candidate genome changes that may modifypathogenicity of the underlying *ABCD1* variant, leading to marked difference in clinical outcomes in these sibs.

**3-Identify epigenetic modifiers using genome-wide methylation profiles of peripheral blood cells (Yrs 1-4; *Kobor, Matthews*)**

DNA and RNA will be simultaneously extracted using a commercial kit (Qiagen). DNA will undergo bisulfite treatment to convert unmethylated cytosine to uracil, leaving methylated Cs protected from conversion. Bisulfite-converted DNA is applied to the Illumina Infinium Methylation EPIC Beadchip array in which the DNAM methylation status of >850,000 CpG sites are examined. The Kobor lab will use their established pipeline to generate the genome-wide methylation profiles of the siblings. Deconvolution methods for array-based methylation profiles from blood[40] will be applied to identify and correct for blood cell-type heterogeneity as a contributor of methylation variability; and multiple test correction will be done using the Benjamini-Hochberg method[41]. In parallel, differentially-methylated regions (DMRs) will be identified using established methods[42,43]; and highly variable DNA methylation sites will be determined using Levene's test for variance. Prioritization of epigenetic differences will be based on the frequency and magnitude of the DNA methylation differences along with overlap of regulatory regions (e.g. enhancers and transcription factor binding sites).

Outcomes: prioritized list of epigenetic differences commonly observed between affected and unaffected individuals as well as prioritized sibling-pair specific DMRs (input for Obj 5).

**4- Lipidomics to identify/validate metabolic markers to complement the genomic/epigenetic data (Yrs 1-4; *Kemp, Wanders*)**

Fasting plasma samples (replicates for 35 discordant sibs, taking treatment status into account) will be analyzed using the AMC lipidomics platform (mass spectrometry with (ultra)-high-resolution/mass accuracy measurement capabilities) which encompasses phospholipids, neutral lipids, gangliosides, sulfatides, sphingolipids, ceramides, and can detect and identify >800 (targeted) discrete and known lipids and

metabolites (this list of known species is continuously growing) and >5000 (untargeted) yet non-classified lipid species/metabolites[44,45]. These data will be used to:

- identify metabolic markers of pathways that are up/down regulated, allowing for in-depth analysis of the epi/genomic data for associated candidate variants/modifiers;
- validate potential modifiers identified by epi/genomics, in genes other than *ABCD1*, using the lipidomic profile as functional read-out (funding for this work has been obtained);
- identify potential dietary or gut microbial factors affecting disease progression. A study involving monozygotic and dizygotic twin pairs that investigated the gut microbiome showed that while there was a high degree of shared microbial genes among sampled individuals, deviations from this core microbiome were associated with different physiological states (Turnbaugh et al. Nature PMID 19043404). Furthermore, microbiota-derived metabolites signal to distant organs in the body, thereby affecting the immune and hormone system and host metabolism (Schroeder et al., Nat Med. 2016 Oct 6;22(10):1079-1089. PMID 27711063). Indeed, studies in mice have shown that the gut microbiota can act as an environmental factor that may promote metabolic diseases by affecting the host energy and lipid metabolism, including energy metabolites and various lipid classes.[46]

Outcomes: lipidomics-based ranking and prioritization of genetic variants as well as dietary/microbial factors hypothesized to serve as modifiers of ALD; functional read-outs to validate genetic modifiers

**5- Integrated bioinformatics of -omics datasets (Yrs 1-4; *Mostafavi, Wasserman, Sinclair*)**

To take advantage of the multiple sets of analyzed and prioritized omics data (genomes, lipidomes, methylomes: *Obj 2-4*) for these 35 well-characterized discordant sib pairs, we will apply an an-house integrated statistical approach based on multi-omics analysis to prioritize potential modifier loci. In particular, all variants that are either unique or of unique genotype (homozygous vs heterozygous) for

each affected sibling will be assigned a *functional* score based on several types of evidence, including genomic annotations encompassing both non-coding variants[38,47] and coding variants[48] evidence of epigenetic effects based on identified DMRs (*Obj2&3);* and evidence of lipidomic *(Obj4)* effects based on the identified metabolic markers of pathways. For identification of (non-coding) regulatory variants, we will use established pipelines, based on quantitative loci analysis (QTL), that have previously mapped regulatory regions to target genes[38,49,50], as well as enhancer-promoter mapping from FANTOM5[51]. DMRs will be associated to nearest target genes within a topologically associating domain[52] or within a maximum 1Mb *cis* distance from the target genes' most proximal Transcription Start Site (TSS)—additional analysis will also investigate assigning DMRs to multiple genes in gene dense regions, and will then utilize QTM mapping to resolve the most likely gene target[53]. We will also use knowledge about gene regulatory networks and pathways to prioritize omics-associated variants. In particular, using the label propagation algorithm developed by PA Mostafavi,[54] evidence from DMRs and differential metabolites target genes will be propagated in gene regulatory networks so that both "direct" and "indirect" network-based evidence will inform the variant functional scores – akin to an approach recently implemented[55] but expanded in scope to also include plausible variants and DMRs in the network construction and analysis.
Outcomes: Identification and prioritization -omics associated variants (functional scores); network-based evidence for variant validation; systems biology algorithms and pipelines

**6- Validation studies using external cohorts and stem cell studies (Yrs 2-4)**
***Independent cohorts (Eichler, Sirrs, van Karnebeek, Braverman, Bernard, Engelen*)** Modifiers will be validated by targeted (epi-) genetic analysis in independent ALD cohorts, ie DNA methylation/Sanger profiling of unrelated ALD patients with the phenotype of interest. These patients (target n=50) will be recruited through collaboration with Vancouver (BCCH & VGH), Montreal (MCH), AMC, ALD Connect and other international ALD Registries.
***Stem cell studies (Heine & Kemp):*** To validate newly identified biomarkers and create new diagnostic measures in neural cell types, we will use induced pluripotent stem cell (iPSC) and CRISPR/Cas9[56] technology to model ALD *in vitro* and test metabolic changes in iPSC-derived neural cell types. PA Heine will reprogram ALD patient cells into iPSCs. If particular genetic variants are not available in the repository (>50 patient lines), CRISPR/Cas9 will be used to genetically edit human control iPSC lines. Patient and genetically-edited iPSC lines will be differentiated into astrocytes according to an adapted protocol[57], and analyzed for known markers of the astrocyte lineage (GFAP, Aquaporin4, GLAST, GLT1, CD44, Ezrin, ALDH1L1) by immunocytochemistry, RT-PCR and western blot. Human iPSC-derived astrocytes can be generated in pure and large amounts, which make them suitable for validation studies. If the identified biomarker is neuronal specific, we will generate neuronal cell types and analyze them accordingly. Functional validation studies will depend on candidate modifiers but will be informed by the associated pathways and the patient's lipidomics profile. The finding of disease modification associated with a *CYP4F2* variant[26] provides an example. Cell lines engineered to contain the variant could be analyzed with targeted lipidomics to verify the predicted alterations in VLCFA omega oxidation. Informative lipid biomarkers from these studies could be analyzed in our validation cohort to confirm *in vivo* alterations in lipid metabolism in individuals with the variant.
Outcomes: a set of brain cell types with the newly identified specific modifier variants, which allow direct functional characterization and validation of the variants at the biochemical/metabolic level.

**PROJECT MANAGEMENT**
*NPA van Karnebeek* is responsible for overall leadership; she is the ideal candidate to lead this research as she is affiliated with both the UBC and the University of Amsterdam. Operationally, she will be assisted by project manager E Lomba, who has 10 years' experience in research administration, HR

and budgets/finance at UBC. With PA Mostafavi at CMMT/UBC, they will work with the Canadian, Dutch and US sites to ensure support and progress. Monthly teleconferences (all sites) will monitor progress, and a central secure platform will be used to communicate and share/track results. Every 3 months the sites will be asked to provide a log of their progress and plans to the NPA. At BCCH, a research coordinator will be responsible for ethics protocols; coordinating study IDs, sample collection and shipments to labs; data streams/analysis; and progress reports. Most sib-pairs are followed at AMC, where a coordinator supervised by Engelen & Kemp will oversee study enrollment, data coordination and sample distribution. In Canada, the UBC and McGill teams will work with Dr. Eichler in ALD-Connect for sample collection for validation studies. **Secure storage & data privacy** Subjects will be assigned de-personalized identifiers to protect privacy. Patient data will be stored in a secure online database in the AMC *(Open Clinica*). All DNA sequence data will be processed on secure computer servers at CFRI/CMMT. Once processed, raw data will be archived to tape (not network accessible). The variant calls and subsequent annotation data will be generated on the secure server and the results delivered to clinical and integrated bioinformatics teams, who maintain them on local secure networks. Epigenomics and lipidomics data are not identifiable. For integrated -omics data analysis, Mostafavi has direct access to the CMMT-stored data; AMC will send lipidomics raw data to her.

**TIMELINE** (*see Gantt chart in Fig 3*) The discordant sib pairs have been identified with ongoing phenotyping. DNA and lipidomics samples have been collected for 25/35 sibs. We are applying for IRB approval, to launch the Project in mid 2017 (when funding opens). Our 4-year timeline is realistic to achieve the milestones and generate the deliverables.

| Year | Year 1 | | | | Year 2 | | | | Year 3 | | | | Year 4 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| QUARTER (3 months) | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| **OBJECTIVE 1: PATIENT ENROLMENT, SAMPLE COLLECTION, PHENOTYPING & FOLLOWUP** | | | | | | | | | | | | | | | | |
| Enrol 35 discordant ALD brothers (*sibling pairs already identified*)<br>• Obtain Ethics approval<br>• Obtain Patient consents<br>• Sample collection (*for 20 sib pairs, samples already available)*<br>• Phenotyping & clinical follow-up | | | | | | | | | | | | | | | | |
| **OBJECTIVE 2: *Identification of genetic (nuclear and mitochondrial) and viral pathogen modifiers of ALD*** | | | | | | | | | | | | | | | | |
| • Bio-informatics pipeline improvements<br>• Sequencing (WGS)<br>• Bioinformatics analysis (*mapped & unmapped reads*)<br>• Sanger re-sequencing | | | | | | | | | | | | | | | | |
| **OBJECTIVE 3: *Identification of epigenetic modifiers of ALD*** | | | | | | | | | | | | | | | | |
| • Methylation arrays<br>• Bioinformatics analysis | | | | | | | | | | | | | | | | |
| **OBJECTIVE 4: *Identification/validation modifiers using lipidomic analyses*** | | | | | | | | | | | | | | | | |
| • MS with (ultra)-high-resolution/mass accuracy measurement capabilities | | | | | | | | | | | | | | | | |
| **OBJECTIVE 5: *Integrated bio-informatics of -omics datasets*** | | | | | | | | | | | | | | | | |
| • Bioinformatics / statistical analysis (*functional networks / regulatory elements*) | | | | | | | | | | | | | | | | |
| **OBJECTIVE 6: *Validation studies*** | | | | | | | | | | | | | | | | |
| Targeted Sanger Analysis ALD cohorts | | | | | | | | | | | | | | | | |
| iPS cells<br>• Generation & characterization of patient genetically-edited iPSC lines<br>• Generation & characterization of iPSC-derived neural cell types<br>• Identification of biomarkers in iPSC-derived astrocytes / neural cell types | | | | | | | | | | | | | | | | |
| **KNOWLEDGE TRANSLATION AND DISSEMINATION** | | | | | | | | | | | | | | | | |
| • Knowledge translation/dissemination (lay, families, scientific/medical public) | | | | | | | | | | | | | | | | |

**Figure 3.** Gantt Chart for our 4-year research project.

## KNOWLEDGE TRANSLATION *(Yrs 1-4)*

**The most tangible outcome** of this multi-omics data analysis is identification of predictive biomarkers. Direct application into clinical practice will be led by our neurometabolic disease specialists (*van Karnebeek*, *Wijburg*, *Bernard*, *Eichler*, *Engelen*) and clinical lab scientists (*Sinclair*, *Wanders*, *Braverman*), whose collective goal is to enhance early diagnosis and intervention to allow affected individuals to reach their full potential. This research and **its immediate translation into care** are urgent now that ALD is being implemented in NBS panels. New York[58] State initiated ALD NBS in 2014; in Feb 2016, Secretary Burwell (US Health & Human Services) issued a formal recommendation to add ALD to the US Recommended Uniform Screening Panel. The Netherlands will start ALD NBS in 2017. In Canada, inclusion of ALD in future panels is under consideration. NBS allows prospective monitoring and early intervention for IEM and other infant-onset diseases. Early diagnosis of boys with ALD is important for early detection of two treatable problems: adrenal insufficiency, to initiate adrenal steroid replacement therapy; and for CALD, to offer allogeneic HSCT. At present, follow-up is identical for all males with ALD, with frequent MRI scans. Predictive biomarkers or risk factors for CALD would enable risk-stratification into low- and high-risk individuals. This will pave the way to precision medicine for ALD (possibly even preventive HSCT long before onset of CALD). If novel therapeutic targets are uncovered, this will serve as catalyst for further research and funding. **Dissemination of new knowledge** will be ongoing, *via* presentations, publications, and social media geared to the lay public (focusing on patients/families), as well as to scientists and health care professionals. Our Team comprises trainees at different levels, thus investing in the next generation of researchers and clinicians, with a focus on translating systems biology data into improved diagnosis and treatment for neurodegenerative diseases such as ALD.

## CHALLENGES & MITIGATION

Technological advances have allowed for the non-targeted approaches that lend significant strength to this proposal; however, the data generated introduces challenges in triaging candidate genes/variants for functional validation. Routes for simplifying the initial datasets include:

- Using discordant sib-pairs allows us to filter variants based on a differential genotype, greatly reducing the number of epi/genomic candidates to be considered.
- The integration of lipidomics data allows for triaging candidate genes based on target pathways, focusing validation studies on genes related to the individual biochemical alterations.
- Deep phenotyping can drive candidate gene selection on an individual basis, realizing the promise of this integrated systems biology approach.

There is also a potential risk of failing to identify strong modifiers. To date, Genome Wide Association-based ALD modifier studies with specific clinical phenotypes (CALD vs AMN) have largely resulted in non-significant findings[8] These studies were largely targeted SNP-based associations, which require very large datasets (not possible with ALD), or strong modifier effects to reach significance. Using discordant sib-pairs and a non-targeted systems biology approach will filter out many false associations allowing candidate gene selection to be guided by individual phenotypic and biochemical markers *vs*. statistical calculations. Similarly, the ability to analyse our datasets and sib-pairs both individually, and as an integrated set, will increase our chances of identifying modifiers that may only be detectable in one dataset or even just an individual sib-pair.

There is a risk that primary modifiers of disease are environmental. Our non-targeted epigenomic, genomic and lipidomic approaches will allow for evaluation of non-genetic contributions to phenotypic discordance. The lipidome, while reflective of primary biochemical pathways, is largely influenced by both dietary and gut microbiotic factors and pathway analysis may reveal dietary or gut microbial targets[59]. For lipidomics analyses, data sets of healthy, aged-matched fasting control samples (via the AMC) will help us identify relevant differences in the patient group. Similarly, human WGS data

contains non-human sequence data with the potential to uncover viral pathogens involved in the initiation of an inflammatory response. Furthermore, blood samples may not prove the 'right' sample type, i.e. not reflective of the central nervous system, for lipidomics and epigenomics, however CNS biopsies are unethical, so these are the best possible. To mitigate this challenge, we will use iPSC-derived neural cell types, and compare variant to control, to characterize and validate any newly identified genetic and lipidomic biomarkers *in vivo*.

We acknowledge our sample size (35 sib pairs) is relatively small; however for a rare disease this represents a unique collection (70 male ALD patients) and thus the opportunity to develop a successful approach to delineation of phenotypic diversity and precise modifiers, and subsequent translation into management and treatment optimization. We will apply for funding to expand the current numbers and research activities. Until our study is performed, we will not be able to perform power calculations as – despite the recent identification of *CYP4F2* by our group-[26] we currently do not have sufficient insight into modifiers for ALD, particularly if there is a strong single modifier or commonly altered gene. Absent such information, a power analysis would be highly speculative. Most of the published genomics studies, which have demonstrated success in revealing key genetic contributors to / candidate modifiers of the phenotypes and outcomes of rare monogenic conditions, work with similarly sized cohorts of patients: *TopBP1* as a novel gene in idiopathic pulmonary arterial hypertension (=12 pts),[63] and *DCTN4* as a modifier of chronic Pseudomonas aeruginosa infection in cystic fibrosis (n=96 pts).[64]

## 2.2 EXPERTISE, EXPERIENCE, AND RESOURCES

Our Team of scientists and clinicians from UBC, McGill, Harvard Medical School, and two Dutch universities is well-placed for success, with complementary expertise and resources including ALD leadership; direct access to patients and samples from 2 international registries; expertise in phenomics, epi-/genomics, metabolomics, iPScells; and a unique ability to interpret large datasets using network approaches.

***NPA-*** *C van Karnebeek* is a clinician-scientist (50% research) in biochemical genetics, pediatrics and applied genomics with a track record of developing novel treatments for neurometabolic diseases and leading major Canadian/international discovery projects. Affiliated with UBC and the University of Amsterdam, she is poised to coordinate the Project from UBC and AMC, working with site leaders to ensure milestones are met (see Project Mngt).

***PAs****: S Mostafavi* (UBC) brings expertise in computational science and statistical approaches for integrating and interpreting genomics data; she will oversee integrated data analysis. *M Engelen* (AMC) pediatric neurologist who follows a large ALD cohort and hosts its biobank will oversee phenomics of ALD siblings and matched controls, data coordination, sample distribution. *S Kemp* (AMC) is a PhD biochemist who heads the ALD research group at AMC; he will oversee the lipidomics study.

**Co-applicants & Collaborators*: at UBC***

- *M Kobor* is an epigenomics expert on the biological factors affecting genome function and gene expression; he will oversee epigenomics analysis.
- *G Sinclair* leads the BC NBS program and method development; he is pivotal to KT efforts.
- *W Wasserman* heads a strong bioinformatics group in applied genomics; he will oversee bioinformatics analysis with focus on regulatory networks.
- *M Tarailo-Graovac* has >10y experience using classic and NGS approaches to identify simple and complex genetic modifiers; she is pivotal to identification of genetic and viral pathogen modifiers.
- *S Sirrs* is Clinic Professor Endocrinology, and Director of the Adult Metabolic Diseases Clinic in Vancouver General Hospital (largest in N-America), following a number of ALD patients
- *A Matthews* is a research associate with expertise in DNA methylation analysis and

bioinformatics analysis of genomic data sets, and special interest in X-linked diseases.

***At McGill University, in the Netherlands & USA***

- *V Heine* (*VUMC/VU*) uses iPSC technology to generate *in vitro* disease models and to test new treatment options for brain disorders; she will validate modifiers and treatments in cells.
- *R Wanders* (*AMC*) & *N Braverman* (*McGill*) are experts in peroxisomal metabolism; pivotal in developing clinically applicable biomarkers for better diagnostics.
- *F Eichler (Harvard), G Bernard (McGill), F Wijburg (AMC)* are experts in new diagnosis and treatment for neurometabolic diseases. As leaders for ALD-Connect and Global leukodystrophy InitiAtive (http://theglia.org/), Eichler & Bernard engage with patients and families. As a member of the Dutch NBS Board, Wijburg provides guidance in NBS and clinical HSCT for lysosomal diseases and ALD.

**ENVIRONMENT**

***Canada: Vancouver's BCCH*** and ***CMMT*** well-established core services for data management, statistics, sequencing, computational facilities, biobanking, and animal care, plus research support services (finance, communications, IT). **McGill** University Health Center, has established a Canadian peroxisome disease center of expertise that involves clinical care, peroxisome function testing and gene panels, and translational research for therapies. ***Netherlands: Amsterdam's Academic Medical Centre*** *has* a 30y history of peroxisomal disease research and of following large cohorts of patients; its advanced Laboratory Genetic Metabolic Diseases is led by Wanders. ***VU Medical Center*** Stem Cell Laboratory led by Heine (www.ipscenter.nl) generates iPSCs from patients with neurodevelopmental disorders, establishing new models of glial defects. ***USA: Boston's Harvard Medical School*** Dr Eichler's research laboratory is renowned for delivering new therapies for monogenetic lipid metabolism disorders such as ALD, and leads RD-connect with access to large patient cohorts.

**IMPACT & FUTURE DIRECTIONS**

***Personalized precision medicine for ALD****:* At present, follow-up is identical for all males with ALD, with frequent MRI scans. Predictive biomarkers or risk factors for CALD would enable risk-stratification into low- and high-risk individuals. This will pave the way to precision medicine for ALD (possibly even preventive HSCT long before onset of CALD), and greatly improve care for male patients with ALD. The burden of yearly MRI scans should no be underestimated. Furthermore, prognosis would improve if HSCT could be initiated before clear MRI lesions are present on MRI. Currently, preventative HSCT is not an option because the associated morbidity and mortality is not negligible. However, if a high-risk group can be identified this possibility can be reconsidered. ***Novel Treatments*****:** Identification of genetic or non-genetic (possibly viral) triggers for the onset of cerebral ALD will have treatment implications. If a specific pathway (eg inflammatory) is identified to play a role in the initiation of cerebral ALD, targeted drug development, in this case immune modulation via high throughput screening techniques, becomes a very real option. This will serve as catalyst for further research and funding. ***Model for other IEMs and other neurodegenerative diseases:*** The generation of diverse -omic datasets in this study will allow us to benchmark systems biology algorithms and pipelines, and thus provide a valuable resource to the broader Bioinformatics community. As such, the algorithms developed here will aid in building sophisticated systems biology approaches to study modifiers for other IEMs for which extreme phenotypic variability poses a treatment challenge (eg Gaucher Disease) as well for rare neurodegenerative diseases in general. ***Canada as leader in the field of phenotypic modifier & systems biology research:*** If funded, this project can further establish Canada's leadership in these fields, offering opportunities for funding to enable systems biology research and continuous improvement of cutting-edge technologies and analysis. This promising field is quickly expanding, in relevance to scientific insight and human health.