

Clinical improvement of DM1 patients reflected by reversal of disease-induced gene expression in blood

Remco T.P. van Cruchten*, Daniël van As*, Jeffrey C. Glennon, Baziel G.M. van Engelen**, Peter A. C. 't Hoen**, the OPTIMISTIC consortium, the ReCognition consortium

* Shared first authors, contributed equally

** Shared senior authors, contributed equally

Author affiliations

Remco van Cruchten, PhD; (1) Center for Molecular and Biomolecular Informatics, Radboud Institute for Molecular Life Sciences, Radboud University Medical Center, Nijmegen, The Netherlands

Remco.vanCruchten@Radboudumc.nl

Daniël van As, MD; (1) Department of Neurology, Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center, Nijmegen, The Netherlands; **(2)** Center for Molecular and Biomolecular Informatics, Radboud Institute for Molecular Life Sciences, Radboud University Medical Center, Nijmegen, The Netherlands

Daniel.vanAs@radboudumc.nl

Jeffrey C. Glennon, PhD; (1) Conway Institute of Biomolecular and Biomedical Research, School of Medicine, University College Dublin, Dublin, Ireland

Jeffrey.Glennon@ucd.ie

Baziel G.M. van Engelen, PhD; (1) Department of Neurology, Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center, Nijmegen, The Netherlands

Baziel.vanEngelen@radboudumc.nl

Peter A.C. 't Hoen, PhD; (1) Center for Molecular and Biomolecular Informatics, Radboud Institute for Molecular Life Sciences, Radboud University Medical Center, Nijmegen, The Netherlands

Peter-Bram.tHoen@radboudumc.nl

Several authors of this publication are members of the Radboudumc Center of Expertise for neuromuscular disorders (Radboud-NMD), Netherlands Neuromuscular Center (NL-NMD) and the European Reference Network for rare neuromuscular diseases (EURO-NMD).

Corresponding author

Professor dr. P.A.C. 't Hoen

Centre for Molecular and Biomolecular Informatics, Radboud University Medical Center, Nijmegen, The Netherlands

Tel: +31 243619763

Email: Peter-Bram.tHoen@radboudumc.nl

Abstract

Myotonic dystrophy type 1 (DM1) is an incurable multisystem disease caused by a CTG-repeat expansion in the DM1 protein kinase (*DMPK*) gene. The OPTIMISTIC clinical trial demonstrated positive and heterogenous effects of cognitive behavioral therapy (CBT) on the capacity for activity and social participations in DM1 patients. Here, we performed mRNA sequencing of full blood for 27 patients of the OPTIMISTIC cohort before and after the CBT intervention. We identified 608 genes for which their expression was significantly associated with the disease causing CTG-repeat expansion, as well as 1176 genes significantly associated with the average clinical response towards the intervention. Remarkably, all 97 genes significantly associated with both returned to more normal levels in patients who benefited most from CBT. This trend was consistent with the difference observed between DM1 patients and controls in an earlier study of blood mRNA expression levels, singling these genes out as candidate biomarkers for therapy response. Together these results highlight the ability to find disease relevant information in full blood of DM1 patients, opening new avenues to monitor therapy effects.

Keywords: Myotonic Dystrophy type 1; Biomarker; RNA-seq; Peripheral blood; Therapeutic Response; Lifestyle Intervention;

1. Introduction

Myotonic dystrophy type 1 (DM1) is a neuromuscular disease with a worldwide average prevalence of around 1 in 8,000 people and a high unmet clinical need [1]. DM1 is considered the most frequently occurring adult-onset form of muscular dystrophy. This degenerative multisystem disease is characterized by a wide range of symptoms including myotonia, muscle weakness and dystrophy, fatigue, apathy, cataracts, obesity and insulin resistance. Next to a severe decrease of life quality, DM1 patients suffer from a reduced life expectancy mostly due to problems with cardiac and respiratory function. Currently no curative therapy exists.

DM1 is caused by expansion of a CTG trinucleotide microsatellite repeat in the 3' UTR of the DM1 protein kinase (*DMPK*) gene [2–4]. Unaffected individuals carry up to approximately 37 CTG triplets in *DMPK*, whilst in DM1 patients this ranges from 50 to even a few thousand repetitions. Depending on the inherited repeat length, DM1 can become manifest at birth or early in life but more frequently becomes apparent in adulthood [1]. In general, the disease manifestation is earlier and more severe with longer repeat expansions. Interruption of the CTG repeat by variants such as CCG or CGG are associated with milder symptoms [5]. The expanded CTG repeat is thought to cause disease mainly via an mRNA gain-of-function mechanism, in which aberrant hairpin structures formed by long CUG repeats are central [6–9]. Directly or indirectly, these hairpin structures dysregulate the function of RNA binding proteins from the muscleblind-like (MBNL) and CUGBP Elav-like (CELF) families, leading to widespread disturbed RNA processing and consequently altered functions of various proteins [10–12]. Although proven to be the disease causing mutation, clinical symptoms of DM1 are only moderately associated with the CTG repeat or the dysregulation of specific proteins which suggests an involvement of other mechanisms in symptom expression [13–16].

While there are many promising therapeutic oligonucleotides, small molecule drugs and gene therapies in the (pre)clinical pipeline for some of the signs and symptoms of DM1, none are expected to reach widespread clinical application soon. Physical training and increasing activity are currently being applied to relieve DM1 symptoms with marked improvements in relatively mildly affected DM1 patients [17,18], which has furthermore shown to induce biochemical responses in DM1 mouse models [19,20].

The to date largest clinical trial in DM1 was OPTIMISTIC: Observational Prolonged Trial In Myotonic dystrophy type 1 to Improve Quality of Life- Standards, a Target Identification Collaboration [18]. The OPTIMISTIC clinical trial included over 250 well-characterized DM1 patients from four centers in Europe, where the effects of cognitive behavioural therapy (CBT) and optional graded exercise therapy were closely monitored over 16 months via more than twenty outcome measures. Notably, the CBT intervention was tailored towards the specific needs of the patient in a shared decision making process between patient and psychotherapist, allowing for a personalized intervention. The trial has shown significant, yet heterogenous improvements for various signs and symptoms, as well as the capacity for social activity and participation in DM1 [21].

Here, we set out to find molecular profiles associated with the disease causing CTG repeat and therapy response based on full blood mRNA sequencing before and after the CBT intervention of 27 carefully selected patients from the OPTIMISTIC cohort. Given the accessibility of peripheral blood, it has increasingly been used for the successful identification of disease biomarkers for a variety of neurological and psychiatric disorders such as Duchenne Muscular Dystrophy, Huntington's disease, Major Depressive Disorder and DM1 [22–25].

Furthermore, the multisystem nature of DM1 is known to be reflected by various laboratory abnormalities of blood samples, supporting the relevance of peripheral blood for the identification of disease relevant information [26]. We analyzed gene expression levels as a function of CTG repeat size (as a proxy for disease load/severity) and of the therapy response. Next, we combined these findings and compared the results to various previously published datasets. We were able to identify 608 genes significantly associated with the CTG repeat and further illustrate that 97 of these genes returned towards more normal expression levels in clinical CBT responders.

2.0 Methods

2.1 Ethical committee approval

The OPTIMISTIC study (NCT02118779) was conducted in accordance with the declaration of Helsinki and approved by the medical-ethical scientific committee for human research at each of the four participating clinical centers. Prior to participation in the OPTIMISTIC study, all participants had provided written informed consent, including a consent for the use of their pseudo anonymized blood, urine and DNA samples [18]. Researchers of this study were not able to link the pseudo anonymized samples to patient identities.

2.2 Sample source and patient selection

Samples and metadata used for this study were all gathered during the OPTIMISTIC clinical trial [18]. At the different time points in the trial blood was drawn and a wide range of clinical outcome measures were recorded. Figure one has been generated to illustrate the heterogeneity in changes across all grouped outcome measures, with annotation of all individual outcome measures in the legend.

Patients were selected to represent a balanced subset of the OPTIMISTIC intervention group (n=128) in order to maximize the generalizability of the study findings. Additionally, patients were selected to represent a continuous uniform distribution of therapy response, as assessed by the primary clinical trial outcome the DM1-Activ-c questionnaire [27]. Finally, we selected for patients that were most completely characterized to facilitate future research. To achieve this, patients were selected for which the DM1-Activ-c questionnaire results were available at each time point (n=104), with less than 20% missing values for other outcome measures (n=81) and without a variant CTG repeat (n=74). Homogeneity of baseline disease severity was accounted for by selecting patients that were within one interquartile range (IQR) of the mean for the baseline variables DM1-Activ-c, 6MWT and CTG repeat length (n=45). One patient was excluded because of polypharmacy (n=44). One patient was excluded because of a drop of 57 points of the DM1-Activ-c score between the baseline and 5 month assessments with a subsequent increase of 55 points between the 5 and 10 month assessments (n=43).

For the final selection we aimed for a continuous uniform distribution of both intervention sites and changes in DM1-Activ-c scores after 10 months. The distribution of the 43 patients over the clinical sites A, B, C and D was: A 12, B 11, C 16, D 4. Therefore, it was decided to select all patients from site D. For the remaining 39 patients, patients were randomly selected as long as each selected patient had a unique change in DM1-Activ-c score. Since only 24 unique delta-DM1-Activ-c scores were left, a second round of random selection was performed allowing each delta-DM1-Activ-c score to be present twice. This process of random selection was repeated until a reasonable site and delta-DM1-Activ-c distribution was

achieved, defined as more than 7 patients for site A, B and C, resulting in the final selection of 30 patients. Due to unavailability of samples and unsuccessful RNA sequencing, three patients were later excluded (n=27). The final selection featured a site distribution of 5 times center A, 8 B, 10 C, 4 D, and 22 unique changes DM1-Activ-c scores, with no change in DM1-Activ-c score being present more than twice.

2.3 RNA-sequencing

Blood drawn during the OPTIMISTIC trial was collected in Tempus tubes and centrally stored at the New Castle MRC Centre for Rare & Neuromuscular Diseases biobank with strict SOPs and temperature control (-80°C). RNA was locally isolated in Nijmegen using the Tempus Spin RNA Isolation Kit (Applied Biosystems/Thermo Fisher Scientific) according to manufacturer's instructions. The concentration and RNA Integrity Number (RIN) was checked using Fragment Analyzer (Thermo Fisher Scientific). The mean RIN value was 8.9 and all were > 7.5. Hemoglobin mRNA was depleted using the GlobinClear kit (Thermo Fisher Scientific). Libraries were prepared using NEBNext Ultra II Directional RNA Library Prep Kit (Illumina) according to manufacturer's instructions for a polyA mRNA workflow using UMI-indexed adapters. The size distribution (between 300 and 500 bp) was confirmed using Fragment Analyzer. 150 bp paired end sequencing was performed with a NovaSeq6000 machine (Illumina) at a library concentration of 1.1 nM, generating >30 M read pairs per sample. All raw sequencing data and associated genotype/phenotype/experimental information is stored in the European Genome-phenome Archive (EGA) under controlled access with Dataset ID EGAS00001005830.

2.4 RNA-sequencing primary data analysis

Adapter sequences and low-quality base calls were removed from fastq files using cutadapt 3.4 via TrimGalore 0.6.6 at no other default parameters than the --paired flag [28]. Trimmed fastq files were mapped to the human genome version hg38.95 using STAR 2.7.0 at default parameters and --outSAMtype BAM SortedByCoordinate [29]. After indexing using samtools [30] at default parameters, PCR-duplicates were removed from the bam files using umi-tools dedup with the flags --spliced-is-unique, --paired and --output-stats (Supplementary Table 1)[31]. Strandedness was verified via RSeQC's infer_experiment [32]. After indexing the deduplicated bam files, reads were counted for overlap with hg38.95 genes via HTSeq with parameters --format bam --order pos and --stranded=yes [33]. EPIC, quantIseq and xCell algorithms were applied to the count tables to verify that the cell type compositions were similar at the two time points [34–36]. GATK HaplotypeCaller and Picard GenotypeConcordance were used to check correct matching of samples from the same patient [37,38].

Splice analysis was performed using rMATS v4.1.0 [39] via the same gtf as for STAR/HTSeq with the parameters and flags: -t paired --readLength 150 --variable-read-length --novelSS --libType fr-firststrand --statoff.

2.5 Comparative DM1 datasets

To validate our findings with previously published datasets, gene counts were extracted from the associated publications [11,12] or were obtained from count tables or raw data as described above from GEO entries GSE85984 and GSE67812 [15,25]. The dataset EV10 from Signorelli et al was obtained for the association between gene expression (logFC) and body-or performance test in Duchenne Muscular Dystrophy patients[22].

To compare DM1 samples to controls a two-sided two-sample Wilcoxon test was performed using *row_wilcoxon_twosample* in *matrixTests* R-package on normalized, log-transformed gene counts [40].

2.6 Statistical analysis

All statistical analyses were carried out in R [41]. For gene expression analysis, firstly genes with low read counts before and after CBT were filtered using *edgeR* *filterByExpr* with *group = before/after CBT* and *min.count = 50* [42]. Following, normalized logCPM values and weights were calculated from the filtered read counts with *Voom* in *Limma* [43].

We first set out to explore differential gene expression before and after the CBT intervention. Gene expression values from *Voom* (in logCPM) were separately modelled using mixed effect models with *before/after CBT* (categorical) as fixed effect and patient identity as random effect [1]. Gene weights were also carried over from *Voom*. This analysis has been implemented using *lme* in the *lme4*-wrapper *lmerTest* [44,45]. *lmerTest* estimates a p-value for the contribution of fixed effects to the model via the Satterthwaite's degrees of freedom method. Fit parameters were extracted with R base *summary* and p-values were FDR corrected via the Benjamini & Hochberg method with *stats* *p.adjust* [46].

$$\text{Gene expression} = \text{CBT}(0/1) + (1|\text{Patient}), \text{ gene weights [1]}$$

In order to study the cohort heterogeneity of gene expression changes, we calculated the logCPM based difference in gene expression between before and after the intervention for each patient for the 560 genes significantly associated with the CBT predictor of [1] (adjusted- $p < 0.05$). Patients and genes with similar expression changes were clustered using the R package *heatmap3* based on the complete linkage method for hierarchical clustering, with gene expression values being centered and scaled per gene [47].

Next, using the same methodology as for the CBT intervention effect, we set out to explore the associations of the different clinical outcome measures and CTG-repeat length with gene expression. For this purpose, we separately modelled gene expression values with either one of the outcome measures or the CTG-repeat length (at trial start, [2]) as fixed effect and patient identity as random effect. The categorical CBT covariate (*before/after*) was included for each fit to correct for differences between the two time points.

$$\text{Gene expression} = \text{CBT}(0/1) + \text{CTG_repeat}^* + (1|\text{Patient}), \text{ gene weights [2]}$$

*Same repeat length fitted for both T0 and T10M

To achieve an overarching measure for CBT response, first the changes for each outcome measure were calculated per patient by subtracting the value after 10 months CBT from that at baseline. Where applicable, outcome measures were multiplied by -1 in order to always associate positive changes with an improved health status. Using R base *scale* [41], the changes per outcome measure were then scaled without centering to account for the different scales of the outcome measures. Finally, for each patient a 'Compound Response' score was calculated based on the mean of all scaled outcome measures. L5ENMO, the mean activity during rest, was a control parameter in OPTIMISTIC and was excluded from this analysis.

Analogous to the methodology described for the outcome measure association analyses, genes associated with overarching clinical response were identified by fitting separate mixed

effect models for each gene with the two fixed effects CBT (categorical) and Compound Response, as well as patient identity as random effect [3]. Notably, the Compound Response variable has only been fitted for gene expression after the CBT intervention (zero at baseline). As such, the Compound Response predictor reflects the difference between the two time points that can be attributed towards therapy responsiveness, while accounting for non-therapy specific differences by including the categorical CBT predictor.

Gene expression = CBT(0/1) + Compound_Response(0/cont.) + (1|Patient), gene weights [3]*
**Compound_Response is zero at T0 and continuous at T10M*

For the splicing analysis, the PSI values for splice exclusion (SE) events were extracted from the rMATS output and fitted in linear models similar to those for gene expression. Splice events were filtered by excluding exons from the analysis with less than three mapping reads and one junction spanning read in at least 14 samples.

The R package ggplot2 was used for representation in volcano- and scatter plots [48]. The R package VennDiagram was used to generate the Venn diagram in Figure 5 [49].

2.7 Gene-set enrichment analysis

Gene-set enrichment analyses have been independently implemented for the gene-sets associated with CBT, CTG-repeat length, Compound Response scores and the genes significantly associated with both CTG-repeat length and Compound Response using gProfiler [50]. For the CBT, CTG-repeat length and Compound Response associated genes, the 500 genes with the lowest nominal p-values were ordered (decreasing) based on their absolute regression coefficients. Subsequent enrichment analyses were implemented using the R client of gProfiler with the parameter `orderd_query = TRUE` against a custom background of 10,292 genes expressed in our samples. Multiple testing correction was based on the default setting 'g_SCS'. We tested for enrichment (one-sided) pathways within the WikiPathway database. The gene-set associated with both CTG-repeat length and Compound Response was based on an FDR threshold of 10% for the respective regression coefficients, resulting in a gene set of 311 genes. A regular, non-order weighted ORA analysis was run for this gene-set with `ordered_quere = FALSE`. Similar to the other analyses, one-sided (enrichment) pathway discovery was based on the WikiPathway database with the default setting 'g_SCS' to correct for multiple testing. For all enrichment analyses only significant pathways (p -adjusted < 0.05) are reported.

Exact scripts and the resulting datasets of the statistical analyses are available via https://github.com/Remcovc/DM1_blood_RNAseq/

3. Results

3.1 Sample selection, procedure and analysis of outcome measures

For the identification of blood biomarkers that are associated with the clinical response to the CBT intervention, we selected 27 patients from the OPTIMISTIC cohort. These patients reflected a uniform continuous distribution of therapy responses as assessed by the primary trial outcome, the DM1-Activ-c questionnaire.

The selected set consisted of 14 females and 13 males aged 19-63 years and represented a range of CTG-repeat lengths (Supplemental Figure S1). All selected patients received CBT, and mRNA-sequencing profiles were obtained at baseline and after 10 months of CBT, the primary endpoint of the OPTIMISTIC trial. These patients featured variable extents of improvement/deterioration after CBT based on the different outcome measures that were collected during the trial (Figure 1).

3.2 Marked changes in gene expression after CBT

We first studied the molecular changes that occurred after the CBT intervention by comparing the mRNA expression levels in blood before and after the CBT intervention. We found that 560 genes were significantly up- or downregulated after CBT (277 genes down, 283 up, Figure 2A). Hierarchical clustering of patients based on the changes in these 560 genes revealed substantial molecular heterogeneity within this 10 months timeframe, which was not in concordance with changes in clinical phenotypes (Figure 2B). The four most significant genes were *GGCX*, *ZNF16*, *SERBP1* and *SLC39A8* (Figure 2C). Biological pathways significantly associated with these genes were limited to an immunological pathway and an electron transport chain in mitochondria (Table 1).

3.3 CTG-repeat length associations reflect molecular dysregulation in blood across different studies

Having a readily accessible fingerprint reflecting the molecular dysregulations associated with DM1 are potentially of high value for both clinical and research settings. To this end, we studied the correlations of blood-based gene expression with the disease causing CTG-repeat length assessed at the start of the trial. The assessments were based on small pool PCR, *Acil* digestion and Southern blot, as opposed to (estimations) of CTG repeat length at birth or disease onset. This was done to minimize potential confounding effects related to the progressive nature of the disease and to assure homogeneity in measurement methodology.

Based on this approach, we identified 608 genes significantly associated with the CTG-repeat at an FDR cutoff of 5% (474 positively, 134 negatively, figure 3A). The four most significant genes *RNF170*, *IRS2*, *NDE1* and *PRIMPOL* showed a clear linear relationship between the CTG-repeat length at baseline and expression values, both before and following the intervention (Figure 3B). Most enriched pathways were related to immunological processes (IL-3, IL-4, IL-5 signaling; chemokine signaling pathway), yet also pathways related to adipogenesis, hepatocyte signaling and AGE/RAGE were discovered. Interestingly, the gene *MMP9* was among several of the CTG-repeat associated pathways.

To validate that these findings are disease relevant, we compared the CTG-repeat length effect size with effect sizes reflecting the differences in gene expression between DM1 and controls in various published datasets (Supplemental Figure S3). We found a strong correlation

with a study that performed mRNA sequencing on DM1 and control blood-samples (Pearson Rho: 0.59, figure S3A [25]). However, similar correlations were not observed for studies profiling DM1 and control tissues other than blood (heart fig S3B [15], brain fig S3C [12], muscle fig S3D [11]). Neither were correlations observed with inferred *MBNL* activity or muscle strength measures from another study (fig S3E and S3F [11]). Interestingly, the CTG-repeat length effect size did correlate well (Pearson Rho: 0.42, Figure S3H) with the effect size of body measurements DMD blood samples, while not correlating with physical test assessments (fig S3G)[22]. In the latter study, comparisons with physical tests and body measures were independently based on the first principal component of a set of different clinical assessments of DMD patients. Thus, while DM1-related molecular dysregulations in blood can be validated in other studies, even from other neuromuscular disorders, they do not necessarily reflect expression dysregulation observed in other tissues.

Since DM1 is known as a splicing disease, we also studied the association of the CTG-repeat length with alternative splicing events. Here, four events in three genes (*RBM39*, *FLNA* and *CTSZ*) reached an FDR threshold of <5%, (Supplemental Figure S4). Given the limited and small effects observed we did not further explore these associations.

3.4 Non-significant associations between gene expression and individual clinical outcome measures

Next we studied phenotype-genotype relationships by estimating the associations of gene expression values with individual clinical outcome measures used in the OPTIMISTIC trial. Here we found virtually no significant associations between gene expression and clinical outcomes. Shown in Supplemental Figure S2 are the associations of gene expression values with the DM1-Activ-c questionnaire. The four genes with the lowest nominal p-values were *SREBF2*, *ZNF283*, *SF3B3* and *GSKIP*.

3.5 Significant associations with average clinical response

To account for the heterogenic changes in individual outcome measures, we calculated a compound CBT response score that reflects the average therapy response of all outcome measures used in OPTIMISTIC (Figure 1). The compound response score can be interpreted as an estimate of an overall difference in capacity between the end and the start of the intervention. We were able to identify 1176 genes significantly (FDR < 0.05) associated with the compound response score (384 positive, 792 negative, Figure 4A). The four most significant hits (*PPP1R9B*, *CSNK1G2*, *PPP6R1*, *FBXO48*) show a clear linear relationship between changes in gene expression during the trial and the Compound Response score (Figure 4B). No enriched pathways were identified for this gene set.

3.6 Clinical improvement linked to reversal of disease-induced gene expression

Since we were able to identify genes significantly associated with both the CTG-repeat length as well as the average clinical response, we were interested in their intersection. Among the significant hits of both analyses, we found an overlap of 97 genes (Figure 5A). To further investigate this relationship, we plotted the effect size of the Compound Response score against the effect size of the CTG repeat length for each expressed gene (Figure 5B). For the 97 genes significantly associated with both predictors a remarkable pattern emerged: genes that were lower expressed in patients with a long CTG repeat showed an increase in expression levels in the patients with a good clinical response and vice versa.

This anticorrelation pattern was confirmed by analyzing an earlier dataset comparing gene expression in DM1 and control blood samples [25]. Here the 97 genes showed a similar association with the DM1 phenotype as has been found with the CTG-length in our study, confirming that the gene expression of patients with a good CBT response changed into the direction of the levels observed in healthy controls (Figure 5C). This remarkable relationship could not be explained by possible confounding between CTG-repeat length and Compound Response, as the Compound Response effect size is only slightly affected by first regressing out the CTG-repeat length effect (Supplemental Figure S5). The four genes most significantly associated with both the CTG-repeat length as well as the Compound Response Score (either FDR < 0.025) were *DNAJB12*, *HDAC5*, *TRIM8* and *ZNF22*. Enriched biological pathways were mostly related to immunological pathways (Chemokine and IL-3 signaling, Microglia Pathogen Phagocytosis Pathway; table 1).

In summary, these results suggest that for a subset of genes significantly associated with the biochemical phenotype caused by the CTG-repeat expansion a reversal of disease-induced gene expression occurred in clinical responders. The association with both molecular dysregulation and clinical response makes this subset of genes highly relevant for the discovery of novel therapeutic targets.

4. Discussion

The purpose of this study was the identification of DM1 specific therapeutic biomarkers in peripheral blood. The multisystem nature of DM1 is known to be reflected by laboratory abnormalities of peripheral blood, making it together with its accessibility a promising tissue for biomarker studies in this disease [26]. Hence, we used blood samples of 27 carefully selected DM1 patients from the OPTIMISTIC cohort to study the associations of gene expression with disease severity as well as the response towards the CBT intervention. The careful selection was done to facilitate generalizability of the study findings, as well as to be able to establish linear associations with therapy response. We identified substantial heterogeneity in molecular expression profile changes after the 10 month CBT-intervention. Furthermore, we identified gene sets that were significantly associated with the disease causing CTG-repeat as well as with the average response towards the CBT intervention across different clinical outcome measures. Most interestingly, an overlap of 97 genes among these latter two gene sets has been identified, showing a clear trend of more normal expression levels in clinical responders. Based on these different gene sets several biological pathways relevant to DM1 have been discovered, as well as specific genes/gene-families with ties to neuro(-muscular) disorders.

The OPTIMISTIC study has shown that DM1 patients significantly improve in their (capacity) for activity and social participation after the CBT intervention [18]. It was furthermore hypothesized that CBT may directly or indirectly improve other biological systems affected by the disease. This hypothesis has been confirmed for muscles of the lower extremity, showing an increase in cross-sectional area as a result of the intervention [51]. Here we set out to further explore this hypothesis by investigating changes on the molecular level. These changes may be the result of increased physical activity, which has been linked to differences in gene expression in previous studies [52], but may also be a more direct effect of the psychotherapeutic CBT intervention [53].

In line with the results of an earlier study, we have illustrated that the clinical response towards the CBT intervention was rather heterogenous [21]. A novel addition to this finding

was the illustration that this heterogeneity extends towards changes in molecular profiles within a 10 month timeframe. Importantly, this heterogeneity could not be explained by changes in cellular composition of the blood samples between the two time points, as similarity of cell type composition has been verified. Additionally, this heterogeneity could not be explained by changes of different outcome measures such as the DM1-Activ-c scale. While the CBT intervention likely played a part in this heterogeneity, the magnitude of this contribution could not be assessed due to the lack of a control group. Other factors, such as ageing or seasonal effects may also have contributed to this finding.

Across the different gene sets identified in this study several of the most significant genes (*SLC39A8*, *IRS2*, *FBXO48*) and one WikiPathway (Transcription factor regulation in adipogenesis) were associated with insulin signaling or more broadly related to metabolism/adipogenesis [54–56]. Dysregulation of insulin signaling has been associated with clinical features of DM1 and is an actively ongoing field of investigation [57]. Aberrant insulin signaling has also been found in other diseases of the nervous system such as depression, with indirect improvements being observed after CBT [58]. Interestingly, the anti-diabetic drug Metformin has been shown to improve the mobility of DM1 patients with effect sizes of the six minute walk test comparable to those observed in the OPTIMISTIC study [59]. With increasing therapeutic interest in this area, our findings suggest that disease relevant insulin signaling can be studied on a molecular level in blood samples, highlighting the utility of peripheral blood in this setting.

Similarly, across most of the gene sets we identified several WikiPathways associated with immunological functions (cell-specific immune response, chemokine signaling pathway, IL-3, 4 and 5 signaling). Whilst this may be in part due to a bias introduced by the profiled tissue, the immune system likely plays a role in the DM1 pathophysiology like for many other chronic diseases [60]. As such, blood sample-based immunology studies may be an interesting field of future investigation.

The intersection of the genes significantly associated with the disease causing CTG-repeat as well as the average CBT response across different outcome measures, revealed a subset of 97 genes. These genes are of particular interest for the identification of therapeutic biomarkers, as they are not only associated with the DM1 phenotype, but also showed normalization of expression levels in clinical responders. Among the genes most significantly associated with both were *HDAC5*, *DNAJB12* and *TRIM8*. In total the subset of 97 genes consisted of two HDACs (histone deacetylases, *HDAC5*, *HDAC7*). HDACs play an important role in transcriptional regulation and compounds that inhibit HDAC enzymes are being studied for their potential effect on a range of human diseases, including neurological disorders [61]. The *DNAJB12* protein is member of the Heat shock protein family, with some evidence supporting positive effects of their induction for muscular dystrophy and other muscle wasting conditions [62]. The TRIM family protein *TRIM72* has been shown to be an essential component of cellular membrane repair in muscles, with evidence supporting some positive effects in mouse models of muscular dystrophy [63]. Authors of the same study suggest the potential of other TRIM family members as potential targets in similar disease states, which may support the further investigation of *TRIM8* in DM1. Although mostly associated with therapy response, *RARA* and *RXRA* were also among the overlapping 97 genes. Stimulating retinoic acid signaling has been linked to muscle regeneration in mouse models via increased proliferation of fibro/adipogenic progenitor cells, highlighting the relevance of this pathway as another potentially DM1

relevant drug target [64]. Taken together, these findings confirm the value of whole blood-based expression profiling for the discovery of therapeutic biomarkers in DM1.

Interestingly, the genes significantly associated with the CTG-repeat showed a moderate correlation with the genes associated with DMD body measurements of an external study. We hypothesize that these body measures are likely also correlated with age, which in turn reflects disease progression. This suggests that some of our results may therefore not be DM1 specific, but rather reflect non-specific molecular dysregulations shared across different (neuromuscular) disorders. This hypothesis is supported by the significant association between the CTG-repeat with *MMP9*, which is known to be a non-specific biomarker that has for instance been linked to cardiac remodeling after myocardial infarction, inflammation and DMD [65,66]. We therefore deem further exploration of shared dysregulations as highly valuable, as this may lead to the discovery of therapeutic targets relevant to a variety of diseases.

Although DM1 is known as alternative splicing disease, only four splice events have been significantly linked to the disease causing CTG-repeat in this study. This may be the result of low *DMPK* expression in blood, which we also believe to be an important contributing factor towards the finding that our CTG-repeat effect size did not correlate with DM1 effect sizes of other tissues from external studies. So, while whole-blood based transcription profiling can identify disease relevant molecular dysregulations, these dysregulations do likely not fully reflect the dysregulations observed in other tissue types. Yet, we found a high correlation of the CTG-repeat effect size with the DM1 phenotype effect size of a different blood based study, as well as with a principal component derived body measure association of a Duchenne Muscular Dystrophy blood based study. While the former validates our findings, the latter suggests the possibility of shared, disease relevant, dysregulations across different neuro(muscular) disorders detectable in peripheral blood. If true, associated pathways might reveal highly interesting targets for drug discovery, as they may have a positive influence on multiple diseases.

4.2 Limitations of this study

To find disease relevant gene-expression in blood, we searched for linear associations with the disease causing CTG-repeat length. While the CTG-repeat length is thought to be the main driver of molecular dysregulation, associations between the progenitor allele length of the CTG repeat with several clinical outcome measures, including DM1-Activ-c and six-minute walk test, have been found to be only weak-moderate [13]. In line with the previously published challenges to directly relate gene expression to clinical phenotypes, we were not able to find significant, direct associations between clinical outcome measures and gene expression. Still, among the most significant genes that were associated with the DM1-Activ-c questionnaire was *GSKIP*, a gene encoding for an inhibitor protein of the known DM1 drug target GSK3- β [67–69]. Given the biological relevance of this finding we deem it likely that the current study design was underpowered to study the association of gene expression with individual clinical outcome measures, especially when taking clinical and molecular heterogeneity into account.

The clinical heterogeneity in therapy response may in part be explained by the personalized nature of the CBT intervention, with therapy foci being tailored towards the needs and wishes of the individual patient. As a consequence, one might expect different outcome measures to be more appropriate for CBT efficacy assessments for different patients. Yet, the identification

of molecular signatures associated with therapy response necessitates the use of the same clinical outcome measure. For this reason, and to average out some of the uncertainty inherently associated with the recording of the different outcome measures, we settled on the use of the compound response score. However, we acknowledge that this combined score is biased by the outcome measures that were used in OPTIMISTIC.

Even though we statistically corrected for non-specific molecular changes between the two time points, the lack of RNA-seq profiles from the OPTIMISTIC control arm makes it difficult to state with certainty that the observed molecular changes are due to the therapy itself. However, this does not discount their value as potential therapeutic targets, as they are, regardless of the mediation factor, significantly associated with improved clinical status. Moreover, for this reason we deemed studying the RNA-seq profiles of the OPTIMISTIC control arm to be less valuable, as these patients did not significantly clinically declined over the 10 month time frame [18].

4.3 Conclusion

Starting from DM1 specific disease determinants, the OPTIMISTIC study has shown that patient tailored CBT can increase the health status of DM1 patients by improving social participation and activity. It was furthermore hypothesized that the CBT intervention positively challenges the biological system, which has already been confirmed by increased cross-sectional area for muscles of the lower extremities. Making use of the clinical heterogeneity in therapy response, we here additionally confirmed disease relevant molecular changes in peripheral blood. Not only do our results highlight the utility of peripheral blood to study the multisystem nature of the disease, but also generated the foundation for an upcoming, multi-omics based drug repurposing study.

5. Acknowledgements

We are grateful to the Newcastle MRC Centre for Rare & Neuromuscular Diseases Biobank, and in particular to dr. Dan Cox, for assistance in the storage and delivery of the patient samples used in this study. We also like to thank Dr. Rick Wansink and Ing. Walther van den Broek for helpful discussions and their work to isolate RNA. Furthermore, we would like to thank dr. Pietro Spitali for the helpful discussions regarding the shared dysregulations observed with Duchenne Muscular Dystrophy.

6. Funding

This study was funded by the European Union's Horizon 2020 research and innovation programme "ERA-NET rare disease research implementing IRDiRC objectives - N° 643578" via the Dutch research funding agency ZON-MW, through the E-Rare Joint Transnational Call JTC 2018 "Translational Research Projects on Rare Diseases" (ReCognitiON project: Recognition and validation of druggable targets from the response to Cognitive Behaviour Therapy in Myotonic Dystrophy type 1 patients from integrated -omics networks). This study was also partially funded by the European Union Seventh Framework Program, under grant agreement no. 305697 (the Observational Prolonged Trial In Myotonic dystrophy type 1 to Improve Quality of Life Standards, a Target Identification Collaboration [OPTIMISTIC] project).

7. Author contributions

Remco van Cruchten: Data Curation, Methodology, Validation, Formal analysis, Writing – Original Draft, Visualization;

Daniël van As: Methodology, Validation, Formal analysis, Writing – Original Draft, Visualization;

Jeffrey C. Glennon: Conceptualization, Funding acquisition, Methodology, Writing – review & editing

Baziel G.M. van Engelen: Conceptualization, Funding acquisition, Methodology, Resources, Supervision, Writing – review & editing

Peter A.C. 't Hoen: Conceptualization, Funding acquisition, Methodology, Supervision, Writing – review & editing

8. Disclosures

R. van Cruchten reports no disclosures relevant to the manuscript;

D. van As reports no disclosures relevant to the manuscript;

J.C. Glennon reports no disclosures relevant to the manuscript;

B. G. M. van Engelen received fees (to the institution) and non-financial support from Fulcrum Therapeutics, Facio Therapies and Arrowhead Pharmaceuticals during the conduct of the study. In addition, he received grant support from the FP7 European Union grand OPTIMISTIC, Marigold Foundation Canada, Prinses Beatrix Spierfonds, Spieren voor Spieren, FSHD Stichting and FSHD Society. He also has an unpaid function as head of the scientific advisory board for Euro-DyMA.

P. A. C. 't Hoen reports no disclosures relevant to the manuscript;

References

- [1] Meola G, Cardani R. Myotonic dystrophies: An update on clinical aspects, genetic, pathology, and molecular pathomechanisms. *Biochim Biophys Acta - Mol Basis Dis* 2015;1852:594–606. <https://doi.org/10.1016/j.bbadis.2014.05.019>.
- [2] Brook JD, McCurrach ME, Harley HG, Buckler AJ, Church D, Aburatani H, et al. Molecular basis of myotonic dystrophy: Expansion of a trinucleotide (CTG) repeat at the 3' end of a transcript encoding a protein kinase family member. *Cell* 1992;68:799–808. [https://doi.org/10.1016/0092-8674\(92\)90154-5](https://doi.org/10.1016/0092-8674(92)90154-5).
- [3] Mahadevan M, Tsilfidis C, Sabourin L, Shutler G, Amemiya C, Jansen G, et al. Myotonic dystrophy mutation: An unstable CTG repeat in the 3' untranslated region of the gene. *Science (80-)* 1992;255:1253–5. <https://doi.org/10.1126/science.1546325>.
- [4] Fu YH, Pizzuti A, Fenwick RG, King J, Rajnarayan S, Dunne PW, et al. An unstable triplet repeat in a gene related to myotonic muscular dystrophy. *Science (80-)* 1992;255:1256–8. <https://doi.org/10.1126/science.1546326>.
- [5] Cumming SA, Hamilton MJ, Robb Y, Gregory H, McWilliam C, Cooper A, et al. De novo repeat interruptions are associated with reduced somatic instability and mild or absent clinical features in myotonic dystrophy type 1. *Eur J Hum Genet* 2018;26:1635–47. <https://doi.org/10.1038/s41431-018-0156-9>.
- [6] Napierała M, Krzyzosiak WJ. CUG repeats present in myotonin kinase RNA form metastable “slippery” hairpins. *J Biol Chem* 1997;272:31079–85. <https://doi.org/10.1074/jbc.272.49.31079>.
- [7] Fardaei M, Rogers MT, Thorpe HM, Larkin K, Hamshere MG, Harper PS, et al. Three proteins, MBNL, MBLL and MBXL, co-localize in vivo with nuclear foci of expanded-repeat transcripts in DM1 and DM2 cells. *Hum Mol Genet* 2002;11:805–14. <https://doi.org/10.1093/hmg/11.7.805>.
- [8] Warf MB, Berglund JA. MBNL binds similar RNA structures in the CUG repeats of myotonic dystrophy and its pre-mRNA substrate cardiac troponin T. *RNA* 2007;13:2238–51. <https://doi.org/10.1261/rna.610607>.
- [9] van Cruchten RTP, Wieringa B, Wansink DG. Expanded CUG repeats in DMPK transcripts adopt diverse hairpin conformations without influencing the structure of the flanking sequences. *RNA* 2019;25:481–95. <https://doi.org/10.1261/rna.068940.118>.
- [10] Nakamori M, Sobczak K, Puwanant A, Welle S, Eichinger K, Pandya S, et al. Splicing biomarkers of disease severity in myotonic dystrophy. *Ann Neurol* 2013;74:862–72. <https://doi.org/10.1002/ana.23992>.
- [11] Wang ET, Treacy D, Eichinger K, Struck A, Estabrook J, Wang TT, et al. Transcriptome alterations in myotonic dystrophy skeletal muscle and heart. *Hum Mol Genet* 2018;28:1312–21. <https://doi.org/10.1093/hmg/ddy432>.
- [12] Otero BA, Poukalov K, Hildebrandt RP, Thornton CA, Jinnai K, Fujimura H, et al. Transcriptome alterations in myotonic dystrophy frontal cortex. *Cell Rep* 2021;34. <https://doi.org/10.1016/j.celrep.2020.108634>.

- [13] Cumming SA, Jimenez-Moreno C, Okkersen K, Wenninger S, Daidj F, Hogarth F, et al. Genetic determinants of disease severity in the myotonic dystrophy type 1 OPTIMISTIC cohort. *Neurology* 2019;93:e995 LP-e1009. <https://doi.org/10.1212/WNL.0000000000008056>.
- [14] Mankodi A, Takahashi MP, Jiang H, Beck CL, Bowers WJ, Moxley RT, et al. Expanded CUG Repeats Trigger Aberrant Splicing of CIC-1 Chloride Channel Pre-mRNA and Hyperexcitability of Skeletal Muscle in Myotonic Dystrophy. *Mol Cell* 2002;10:35–44. [https://doi.org/https://doi.org/10.1016/S1097-2765\(02\)00563-4](https://doi.org/https://doi.org/10.1016/S1097-2765(02)00563-4).
- [15] Freyermuth F, Rau F, Kokunai Y, Linke T, Sellier C, Nakamori M, et al. Splicing misregulation of SCN5A contributes to cardiac-conduction delay and heart arrhythmia in myotonic dystrophy. *Nat Commun* 2016;7:11067. <https://doi.org/10.1038/ncomms11067>.
- [16] Savkur RS, Philips A V., Cooper TA. Aberrant regulation of insulin receptor alternative splicing is associated with insulin resistance in myotonic dystrophy. *Nat Genet* 2001;29:40–7. <https://doi.org/10.1038/ng704>.
- [17] Roussel M-P, Morin M, Gagnon C, Duchesne E. What is known about the effects of exercise or training to reduce skeletal muscle impairments of patients with myotonic dystrophy type 1? A scoping review. *BMC Musculoskelet Disord* 2019;20:101. <https://doi.org/10.1186/s12891-019-2458-7>.
- [18] Okkersen K, Jimenez-Moreno C, Wenninger S, Daidj F, Glennon J, Cumming S, et al. Cognitive behavioural therapy with optional graded exercise therapy in patients with severe fatigue with myotonic dystrophy type 1: a multicentre, single-blind, randomised trial. *Lancet Neurol* 2018;17:671–80. [https://doi.org/10.1016/S1474-4422\(18\)30203-5](https://doi.org/10.1016/S1474-4422(18)30203-5).
- [19] Manta A, Stouth DW, Xhuti D, Chi L, Rebalka IA, Kalmar JM, et al. Chronic exercise mitigates disease mechanisms and improves muscle function in myotonic dystrophy type 1 mice. *J Physiol* 2019;597:1361–81. <https://doi.org/https://doi.org/10.1113/JP277123>.
- [20] Hu N, Kim E, Antoury L, Li J, González-Pérez P, Rutkove SB, et al. Antisense oligonucleotide and adjuvant exercise therapy reverse fatigue in old mice with myotonic dystrophy. *Mol Ther Nucleic Acids* 2020;23:393–405. <https://doi.org/10.1016/j.omtn.2020.11.014>.
- [21] van As D, Okkersen K, Bassez G, Schoser B, Lochmüller H, Glennon JC, et al. Clinical Outcome Evaluations and CBT Response Prediction in Myotonic Dystrophy. *J Neuromuscul Dis* 2021;1–16. <https://doi.org/10.3233/jnd-210634>.
- [22] Signorelli M, Ebrahimipoor M, Veth O, Hettne K, Verwey N, García-Rodríguez R, et al. Peripheral blood transcriptome profiling enables monitoring disease progression in dystrophic mice and patients. *EMBO Mol Med* 2021;13:e13328. <https://doi.org/https://doi.org/10.15252/emmm.202013328>.
- [23] Byrne LM, Rodrigues FB, Blennow K, Durr A, Leavitt BR, Roos RAC, et al. Neurofilament light protein in blood as a potential biomarker of neurodegeneration in Huntington’s disease: a retrospective cohort analysis. *Lancet Neurol* 2017;16:601–9.

[https://doi.org/10.1016/S1474-4422\(17\)30124-2](https://doi.org/10.1016/S1474-4422(17)30124-2).

- [24] Spijker S, Van Zanten JS, De Jong S, Penninx BWJH, Van Dyck R, Zitman FG, et al. Stimulated Gene Expression Profiles as a Blood Marker of Major Depressive Disorder 2010. <https://doi.org/10.1016/j.biopsych.2010.03.017>.
- [25] Sznajder LJ, Scotti MM, Shin J, Taylor K, Ivankovic F, Nutter CA, et al. Loss of MBNL1 induces RNA misprocessing in the thymus and peripheral blood. *Nat Commun* 2020;11:2022. <https://doi.org/10.1038/s41467-020-15962-x>.
- [26] Heatwole CR, Miller J, Martens B, Moxley RT. Laboratory abnormalities in ambulatory patients with myotonic dystrophy type 1. *Arch Neurol* 2006;63:1149–53. <https://doi.org/10.1001/archneur.63.8.1149>.
- [27] Hermans MCE, Hoeijmakers JGJ, Faber CG, Merkies ISJ. Reconstructing the Rasch-built myotonic dystrophy type 1 activity and participation scale. *PLoS One* 2015;10. <https://doi.org/10.1371/journal.pone.0139944>.
- [28] Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnetJournal*; Vol 17, No 1 Next Gener Seq Data Anal - 1014806/Ej171200 2011.
- [29] Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 2013;29:15–21. <https://doi.org/10.1093/bioinformatics/bts635>.
- [30] Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009;25:2078–9. <https://doi.org/10.1093/bioinformatics/btp352>.
- [31] Smith T, Heger A, Sudbery I. UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome Res* 2017;27:491–9. <https://doi.org/10.1101/gr.209601.116>.
- [32] Wang L, Wang S, Li W. RSeQC: quality control of RNA-seq experiments. *Bioinformatics* 2012;28:2184–5. <https://doi.org/10.1093/bioinformatics/bts356>.
- [33] Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* 2015;31:166–9. <https://doi.org/10.1093/bioinformatics/btu638>.
- [34] Racle J, de Jonge K, Baumgaertner P, Speiser DE, Gfeller D. Simultaneous enumeration of cancer and immune cell types from bulk tumor gene expression data. *Elife* 2017;6:e26476. <https://doi.org/10.7554/eLife.26476>.
- [35] Finotello F, Mayer C, Plattner C, Laschober G, Rieder D, Hackl H, et al. Molecular and pharmacological modulators of the tumor immune contexture revealed by deconvolution of RNA-seq data. *Genome Med* 2019;11:34. <https://doi.org/10.1186/s13073-019-0638-6>.
- [36] Aran D, Hu Z, Butte AJ. xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol* 2017;18:220. <https://doi.org/10.1186/s13059-017-1349-1>.
- [37] Poplin R, Ruano-Rubio V, DePristo MA, Fennell TJ, Carneiro MO, Van der Auwera GA,

- et al. Scaling accurate genetic variant discovery to tens of thousands of samples. *BioRxiv* 2018:201178. <https://doi.org/10.1101/201178>.
- [38] Institute B. Picard toolkit. Broad Institute, GitHub Repos 2019.
- [39] Shen S, Park JW, Lu ZX, Lin L, Henry MD, Wu YN, et al. rMATS: Robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc Natl Acad Sci U S A* 2014;111:E5593–601. <https://doi.org/10.1073/pnas.1419161111>.
- [40] Schwender H, Müller T. Computing thousands of test statistics simultaneously in R. *Stat Comput Graph* 2007;18.
- [41] Team RC. R: A Language and Environment for Statistical Computing 2019:<https://www.R-project.org>.
- [42] Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 2010;26:139–40. <https://doi.org/10.1093/bioinformatics/btp616>.
- [43] Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 2015;43:e47–e47. <https://doi.org/10.1093/nar/gkv007>.
- [44] Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using lme4. *J Stat Software; Vol 1, Issue 1* 2015. <https://doi.org/10.18637/jss.v067.i01>.
- [45] Kuznetsova A, Brockhoff PB, Christensen RHB. lmerTest Package: Tests in Linear Mixed Effects Models. *J Stat Software; Vol 1, Issue 13* 2017. <https://doi.org/10.18637/jss.v082.i13>.
- [46] Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J R Stat Soc Ser B* 1995;57:289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>.
- [47] Zhao S, Guo Y, Sheng Q, Shyr Y. Heatmap3: an improved heatmap package with more powerful and convenient features. *BMC Bioinforma* 2014 1510 2014;15:1–2. <https://doi.org/10.1186/1471-2105-15-S10-P16>.
- [48] Wickham H. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York; 2016.
- [49] Chen H, Boutros PC. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinformatics* 2011;12:35. <https://doi.org/10.1186/1471-2105-12-35>.
- [50] Raudvere U, Kolberg L, Kuzmin I, Arak T, Adler P, Peterson H, et al. G:Profiler: A web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res* 2019;47:W191–8. <https://doi.org/10.1093/NAR/GKZ369>.
- [51] Heskamp L, Okkersen K, van Nimwegen M V., Ploegmakers MJ, Bassez G, Deux JF, et al. Quantitative muscle MRI depicts increased muscle mass after a behavioral change in myotonic dystrophy type 1. *Radiology* 2020;297:132–42. <https://doi.org/10.1148/radiol.2020192518>.

- [52] Grazioli E, Dimauro I, Mercatelli N, Wang G, Pitsiladis Y, Di Luigi L, et al. Physical activity in the prevention of human diseases: role of epigenetic modifications. *BMC Genomics* 2017 188 2017;18:111–23. <https://doi.org/10.1186/S12864-017-4193-5>.
- [53] Jiménez JP, Botto A, Herrera L, Leighton C, Rossi JL, Quevedo Y, et al. Psychotherapy and Genetic Neuroscience: An Emerging Dialog. *Front Genet* 2018;0:257. <https://doi.org/10.3389/FGENE.2018.00257>.
- [54] Li D, Achkar J-P, Haritunians T, Jacobs JP, Hui KY, Amato M D', et al. A pleiotropic missense variant in SLC39A8 is associated with Crohn's disease and human gut microbiome composition HHS Public Access. *Ashwin Ananthakrishnan n.d.*;8:21. <https://doi.org/10.1053/j.gastro.2016.06.051>.
- [55] Brady MJ. IRS2 takes center stage in the development of type 2 diabetes. *J Clin Invest* 2004;114:886. <https://doi.org/10.1172/JCI23108>.
- [56] Liu Y, Jurczak MJ, Lear TB, Lin B, Larsen MB, Kennerdell JR, et al. A Fbxo48 inhibitor prevents pAMPK α degradation and ameliorates insulin resistance. *Nat Chem Biol* 2021 173 2021;17:298–306. <https://doi.org/10.1038/s41589-020-00723-0>.
- [57] Nieuwenhuis S, Okkersen K, Widomska J, Blom P, 't Hoen PAC, van Engelen B, et al. Insulin Signaling as a Key Moderator in Myotonic Dystrophy Type 1. *Front Neurol* 2019;0:1229. <https://doi.org/10.3389/FNEUR.2019.01229>.
- [58] Gulley LD, Shomaker LB, Kelly NR, Chen KY, Stice E, Olsen CH, et al. Indirect Effects of a Cognitive-Behavioral Intervention on Adolescent Weight and Insulin Resistance Through Decreasing Depression in a Randomized Controlled Trial. *J Pediatr Psychol* 2019;44:1163–73. <https://doi.org/10.1093/JPEPSY/JSZ064>.
- [59] Bassez G, Audureau E, Hogrel JY, Arrouasse R, Baghdoyan S, Bhugaloo H, et al. Improved mobility with metformin in patients with myotonic dystrophy type 1: A randomized controlled trial. *Brain* 2018;141:2855–65. <https://doi.org/10.1093/brain/awy231>.
- [60] Zhong J, Shi G. Editorial: Regulation of inflammation in chronic disease. *Front Immunol* 2019;10:737. <https://doi.org/10.3389/FIMMU.2019.00737/BIBTEX>.
- [61] Gottesfeld JM, Pandolfo M. Development of histone deacetylase inhibitors as therapeutics for neurological disease. *Future Neurol* 2009;4:775. <https://doi.org/10.2217/FNL.09.55>.
- [62] Thakur SS, Swiderski K, Ryall JG, Lynch GS. Therapeutic potential of heat shock protein induction for muscular dystrophy and other muscle wasting conditions. *Philos Trans R Soc B Biol Sci* 2018;373. <https://doi.org/10.1098/RSTB.2016.0528>.
- [63] Alloush J, Weisleder N. TRIM proteins in therapeutic membrane repair of Muscular dystrophy. *JAMA Neurol* 2013;70:928–31. <https://doi.org/10.1001/jamaneurol.2013.469>.
- [64] Zhao L, Son JS, Wang B, Tian Q, Chen Y, Liu X, et al. Retinoic acid signalling in fibro/adipogenic progenitors robustly enhances muscle regeneration. *EBioMedicine* 2020;60. <https://doi.org/10.1016/j.ebiom.2020.103020>.

- [65] Halade G V., Jin YF, Lindsey ML. Matrix Metalloproteinase (MMP)-9: a proximal biomarker for cardiac remodeling and a distal biomarker for inflammation. *Pharmacol Ther* 2013;139:32. <https://doi.org/10.1016/J.PHARMTHERA.2013.03.009>.
- [66] Lourbakos A, Yau N, De Bruijn P, Hiller M, Kozaczynska K, Jean-Baptiste R, et al. Evaluation of serum MMP-9 as predictive biomarker for antisense therapy in Duchenne. *Sci Reports* 2017 71 2017;7:1–10. <https://doi.org/10.1038/s41598-017-17982-y>.
- [67] Wei C, Jones K, Timchenko NA, Timchenko L. GSK3 β is a new therapeutic target for myotonic dystrophy type 1. *Rare Dis* 2013;1:e26555. <https://doi.org/10.4161/RDIS.26555>.
- [68] Wei C, Stock L, Valanejad L, Zalewski ZA, Karns R, Puymirat J, et al. Correction of GSK3 β at young age prevents muscle pathology in mice with myotonic dystrophy type 1. *FASEB J* 2018;32:2073–85. <https://doi.org/10.1096/fj.201700700R>.
- [69] Wang M, Weng W-C, Stock L, Lindquist D, Martinez A, Gourdon G, et al. Correction of Glycogen Synthase Kinase 3 β in Myotonic Dystrophy 1 Reduces the Mutant RNA and Improves Postnatal Survival of DMSXL Mice. *Mol Cell Biol* 2019;39:e00155-19. <https://doi.org/10.1128/MCB.00155-19>.

Table 1: Gene set enrichment analysis

query	Adjusted p-value	Term size	Intersection size	Term ID	Term Name	Intersection with WikiPathway
CBT_ORA	0.014	49	5	WP:WP5039	SARS-CoV-2 Innate Immunity Evasion and Cell-specific immune response	<i>CXCL1, PPBP, PF4, CXCR2, CXCL5</i>
CBT_ORA	0.037	91	8	WP:WP111	Electron Transport Chain (OXPHOS system in mitochondria)	<i>MT-ND3, ATP5F1E, MT-ND1, MT-ATP6, MT-CO2, MT-CO3, MT-ND4, MT-CO1</i>
CTG_ORA	<0.001	46	10	WP:WP286	IL-3 Signaling Pathway	<i>FOS, CSF2RB, GAB2, HCK, LYN, MAPK3, RAF1, STAT3, STAT5B, PIK3CD</i>
CTG_ORA	0.001	49	10	WP:WP395	IL-4 Signaling Pathway	<i>FOS, IRS2, CEBPB, GAB2, NFIL3, MAPK3, STAT6, STAT3, STAT5B, PIK3CD</i>
CTG_ORA	0.002	26	3	WP:WP4564	Neural Crest Cell Migration during Development	<i>MMP9, EPHB4, FOS</i>
CTG_ORA	0.003	28	3	WP:WP4565	Neural Crest Cell Migration in Cancer	<i>MMP9, EPHB4, FOS</i>
CTG_ORA	0.004	29	7	WP:WP3972	PDGFR-beta pathway	<i>FOS, MAPK3, RAF1, MAP2K4, STAT6, STAT3, STAT5B</i>
CTG_ORA	0.006	29	2	WP:WP4808	Endochondral Ossification with Skeletal Dysplasias	<i>MMP9, ALPL</i>
CTG_ORA	0.006	29	2	WP:WP474	Endochondral Ossification	<i>MMP9, ALPL</i>
CTG_ORA	0.006	56	2	WP:WP2324	AGE/RAGE pathway	<i>MMP9, ALPL</i>
CTG_ORA	0.008	48	10	WP:WP304	Kit receptor signaling pathway	<i>FOS, GAB2, SH2B2, LYN, MAPK3, RAF1, STAT3, STAT5B, CRK, RPS6KA1</i>
CTG_ORA	0.011	37	8	WP:WP127	IL-5 Signaling Pathway	<i>FOS, CSF2RB, LYN, MAPK3, RAF1, STAT3, STAT5B, RPS6KA1</i>
CTG_ORA	0.013	31	7	WP:WP313	Signaling of Hepatocyte Growth Factor Receptor	<i>FOS, PTEN, MAPK3, RAF1, STAT3, CRK, PXN</i>
CTG_ORA	0.015	110	16	WP:WP3929	Chemokine signaling pathway	<i>CXCR2, TIAM2, HCK, NCF1, LYN, ARRB2, GNB2, PREX1, MAPK3, RAF1, STAT3, STAT5B, CRK, WAS, PXN, PIK3CD</i>
CTG_ORA	0.031	15	3	WP:WP3599	Transcription factor regulation in adipogenesis	<i>IRS2, CEBPD, CEBPB</i>
CTG_CR_ORA	0.002	110	13	WP:WP3929	Chemokine signaling pathway	<i>WAS, PXN, HCK, MAPK3, PREX1, ARRB2, PIK3R5, VAV1, NCF1, CXCL16, PIK3CD, GNB2, GRB2</i>
CTG_CR_ORA	0.006	46	8	WP:WP286	IL-3 Signaling Pathway	<i>GAB2, CSF2RB, HCK, MAPK3, PTPN6, VAV1, PIK3CD, GRB2</i>
CTG_CR_ORA	0.009	48	8	WP:WP304	Kit receptor signaling pathway	<i>GAB2, MAPK3, PTPN6, RPS6KA1, VAV1, SH2B2, JUNB, GRB2</i>
CTG_CR_ORA	0.016	115	12	WP:WP306	Focal Adhesion	<i>ACTN1, PXN, HCK, MAPK3, CCND2, VASP, TLN1, VAV1, ZYX, ITGA5, PIK3CD, GRB2</i>
CTG_CR_ORA	0.032	31	6	WP:WP3937	Microglia Pathogen Phagocytosis Pathway	<i>HCK, PTPN6, VAV1, NCF1, ITGB2, PIK3CD</i>

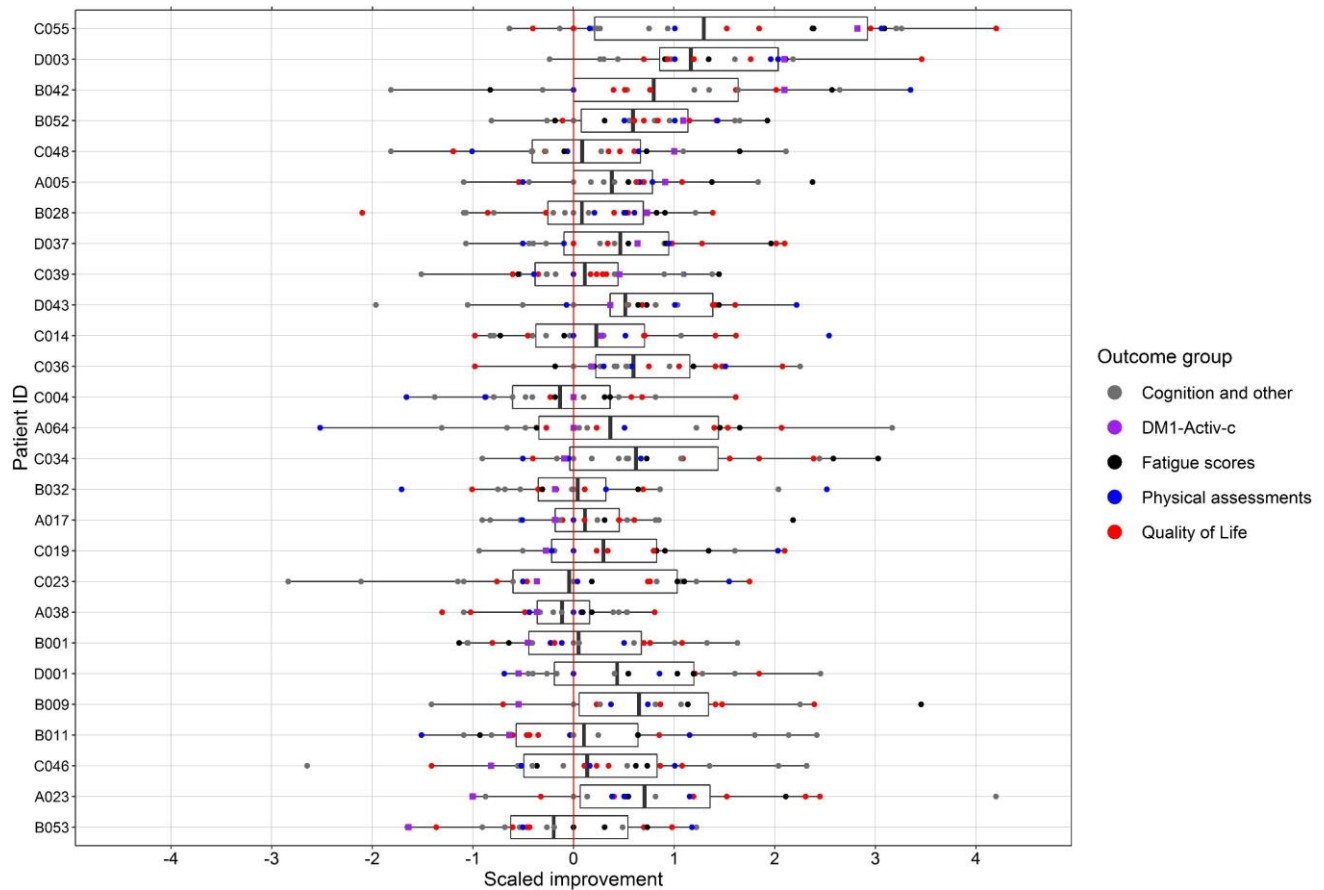


Figure 1: Distribution of changes in outcome measures per patient

Per outcome measure, changes between baseline and 10 months of CBT were scaled by the root mean square. Additionally, for some outcome measures a sign-adjustment was performed so that an increased score is always associated with improved health status. All outcome measures are shown per patient, where patients were ordered along the y-axis by their change in DM1-Activ-c scores (purple squares). Boxes enclose the 25th to 75th percentiles, divided by a thick line that represents the mean compound response score. The whiskers represent the lowest/highest value no further than 1.5 times the interquartile range.

Quality of life (red): Myotonic Dystrophy Health Index, Individualized Neuromuscular Quality of Life Questionnaire, Adult Social Behavioural Questionnaire, Illness Management Questionnaire, Checklist individual strength – Subscale activity.

Physical assessments (blue): Six Minute Walk Test, BORG Scale, Accelerometry measures

Fatigue scores (black): Fatigue and Daytime Sleepiness Scale, Checklist Individual Strength – Subscale fatigue, Jacobsen Fatigue Catastrophizing Scale

Cognition and other (gray): Trail Making Test, Stroop Color-Word Interference Test, McGill Pain Questionnaire, Beck Depression Inventory – Fast Screen, Social Support – Discrepancies and Negative Interactions, Apathy Evaluation Scale – Clinical Version, Self-Efficacy Scale 28

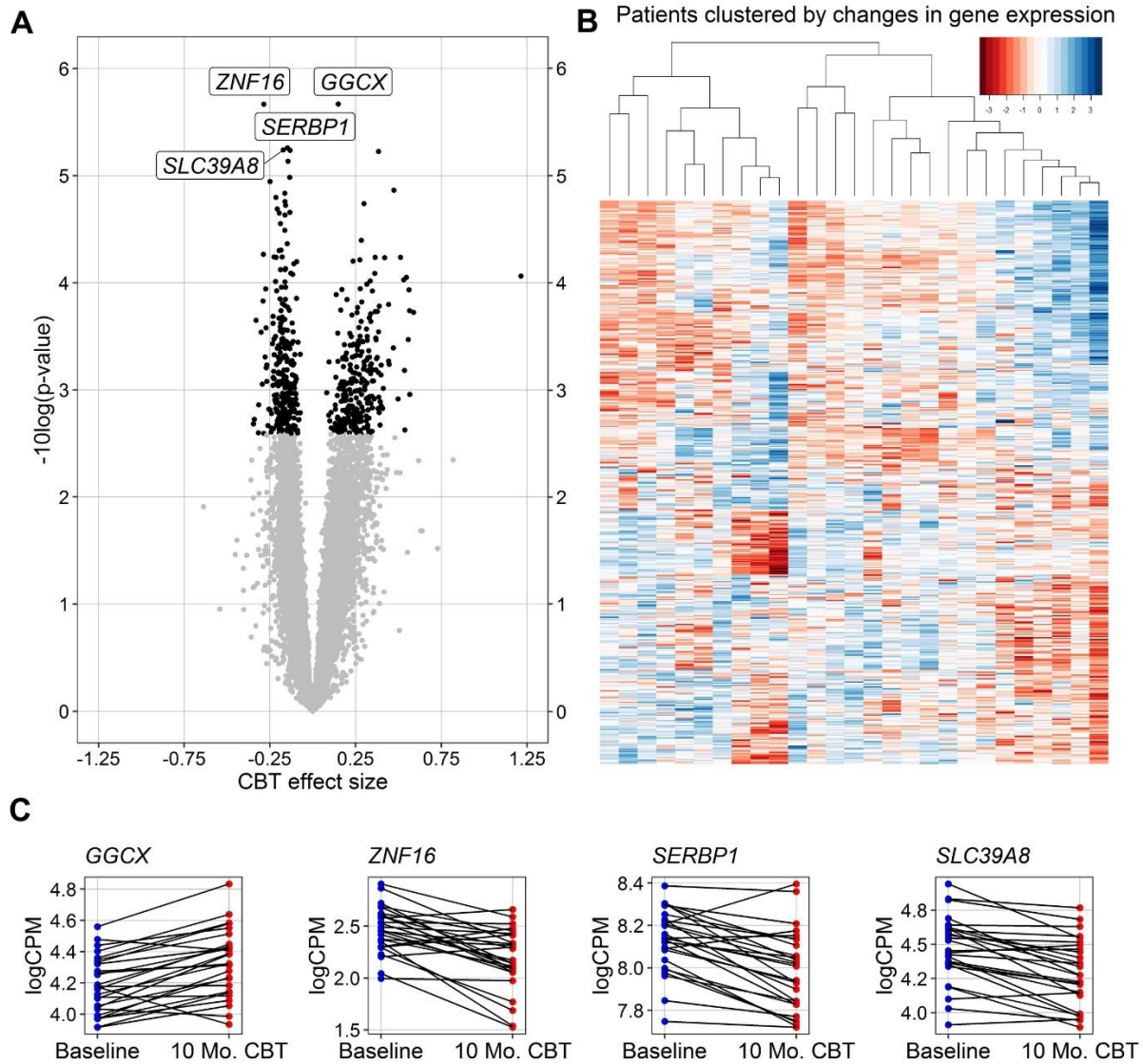


Figure 2: Changes in gene expression after cognitive behavioural therapy

A linear mixed effect model was fitted for each gene, estimating the fixed effect of CBT while accounting for (random) effects of the individual. The p-values for the fixed effect were estimated via Satterthwaite's degrees of freedom method and FDR-corrected. A) Volcano plot of significance ($-\log_{10}$ of the nominal p-value) and the effect size for changed expression after 10 months of CBT. Genes for which the effect size of CBT is significant (FDR < 0.05) are visualized in black. B) Heatmap of changes in normalized logCPM values between the baseline and the 10 months assessment for the 560 genes significantly associated with the CBT effect size. Patients and genes were clustered based on the complete linkage method for hierarchical clustering, values were centered and scaled per gene. C) Expression values (logCPM) at baseline (blue) and after CBT (red) of the four genes with the lowest nominal p-values from panel A.

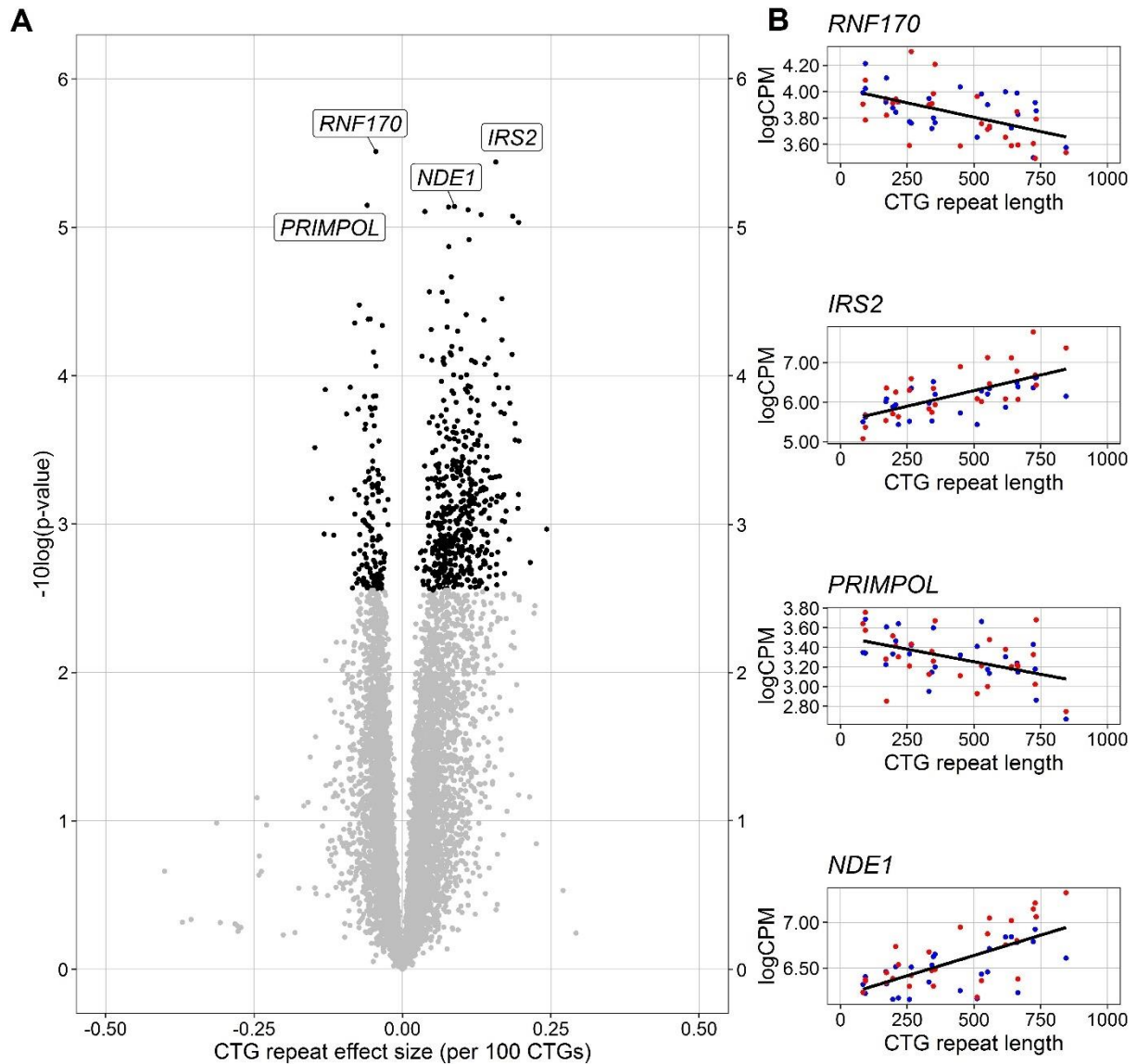


Figure 3: Gene expression levels associated with CTG-repeat length

For each gene a mixed effect model was fitted with before/after CBT and CTG-repeat length as fixed effects, while accounting for (random) effects of the individual. The p-values for the fixed effects were estimated via Satterthwaite's freedom method and FDR corrected. A) Volcano plot of significance ($-\log_{10}$ of the nominal p-value) and effect size of the CTG-repeat length (per 100 CTGs) on gene expression. Genes for which the effect of CTG-repeat is significant (FDR < 0.05) are visualized in black. B) For the four genes with the lowest nominal p-values from panel A, the gene expression values (logCPM) are plotted against the CTG-repeat length. Blue dots represent baseline expression values, red dots expression values after CBT. The regression line is fitted over all values, independent of the timepoint.

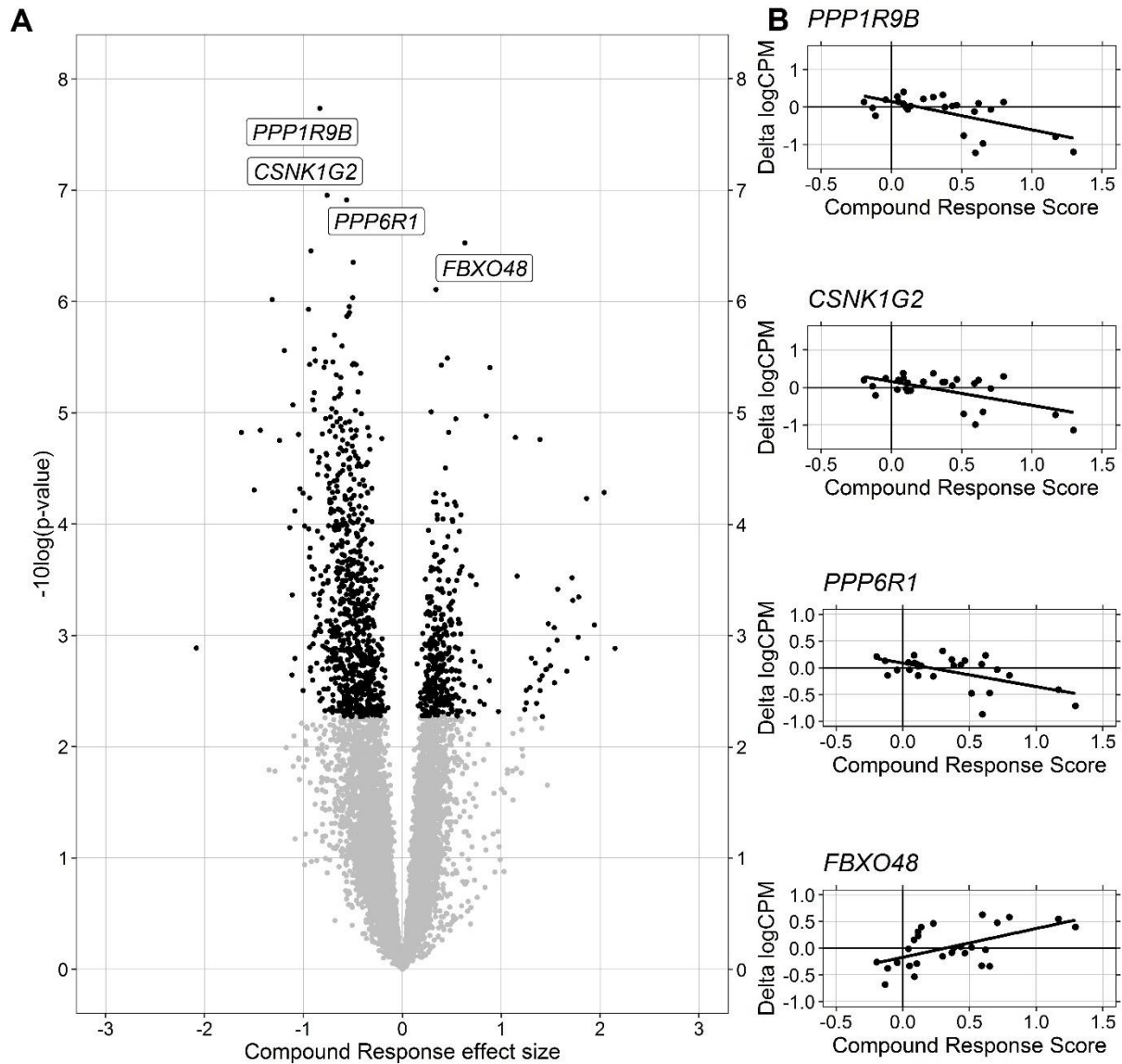


Figure 4: Gene expression association with Compound Response scores

For each gene, a mixed effect model was fitted with before/after CBT and Compound Response Scores as fixed effects, while accounting for (random) effects of the individual. Compound Response Scores were fitted for gene expression values after CBT and set to be zero at baseline, the effect size of this covariate therefore expresses changes in gene expression compared to the baseline values that are associated with clinical response. The p-values for the fixed effects were estimated via Satterthwaite's freedom method and FDR corrected. A) Volcano plot of significance ($-\log_{10}$ of the nominal p-value) and the effect size of the Compound Response Score on gene expression. Genes for which the effect size of Compound Response is significant ($FDR < 0.05$) are visualized in black. B) For the four genes with the lowest nominal p-values from panel A, the changes in gene expression (delta logCPM after-before CBT) are plotted against the Compound Response scores.

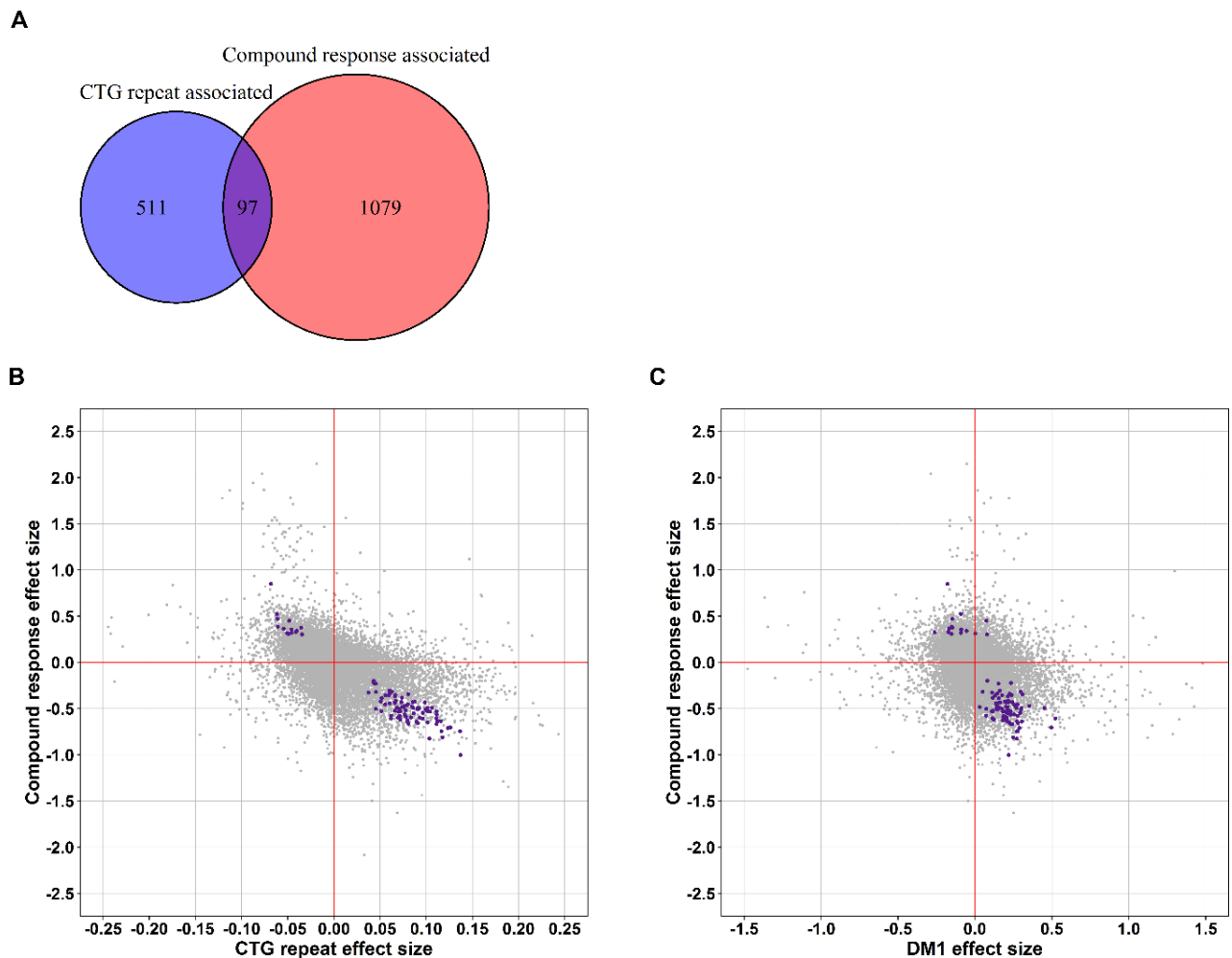


Figure 5: Clinical improvement is linked to normalization of expression of CTG-repeat associated genes

Linear mixed effect models were fitted for each expressed gene, with CBT as fixed effect, patient as random effect, and either CTG repeat or Compound Response as predictor. P-values for the regression coefficients were estimated via Satterthwaite's degrees of freedom and considered significant for values smaller than 0.05 after FDR correction. Furthermore, differences in gene expression of blood samples from DM1 patients and controls were calculated based on an external study using a Wilcoxon signed-rank test on normalized, log-transformed gene counts. A) Venn-diagram illustrating the number of significant genes associated with CTG repeat length and Compound Response, as well as their overlap (disease relevant changes). B) For all expressed genes, the regression coefficients of the Compound Response scores are plotted against the regression coefficients of the CTG repeat lengths. For illustrative purposes the regression coefficients of the CTG repeat have been multiplied by 100. Furthermore, the x-axis has been scaled between -0.25 and 0.25, removing 12 outliers from the figure. Colored in purple are the genes for which both regression coefficients were significant (FDR < 0.05). C) For all expressed genes, the Compound Response effect size is plotted against the DM1 effect size based on an external study comparing blood expression profiles from DM1 patients and controls [25]. For illustrative purposes the x-axis has been scaled between -1.5 and 1.5, removing 6 outliers from the figure.