

RESEARCH

Open Access



# Abnormal molecular signatures of inflammation, energy metabolism, and vesicle biology in human Huntington disease peripheral tissues

Andreas Neueder<sup>1</sup>, Kerstin Kojer<sup>1</sup>, Tanja Hering<sup>1</sup>, Daniel J. Lavery<sup>2,3</sup>, Jian Chen<sup>2</sup>, Nathalie Birth<sup>1</sup>, Jaqueline Hallitsch<sup>1</sup>, Sonja Trautmann<sup>1</sup>, Jennifer Parker<sup>4</sup>, Michael Flower<sup>4</sup>, Huma Sethi<sup>4</sup>, Salman Haider<sup>4</sup>, Jong-Min Lee<sup>5,6</sup>, Sarah J. Tabrizi<sup>4</sup> and Michael Orth<sup>1,7,8\*</sup> 

\*Correspondence:  
michael.orth@siloh.ch

<sup>1</sup> Department of Neurology, Ulm University, 89081 Ulm, Germany

<sup>2</sup> CHDI Foundation, Princeton, NJ 08540, USA

<sup>3</sup> Loulou Foundation, Orphan Disease Center, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA

<sup>4</sup> UCL Huntington's Disease Centre, UCL Queen Square Institute of Neurology and National Hospital for Neurology and Neurosurgery, Queen Square, London WC1N 3BG, UK

<sup>5</sup> Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA 02114, USA

<sup>6</sup> Department of Neurology, Harvard Medical School, Boston, MA 02115, USA

<sup>7</sup> Swiss Huntington Centre, Neurozentrum, Siloah AG, Worbstr. 312, 3073 Gümligen, Switzerland

<sup>8</sup> University Hospital of Old Age Psychiatry and Psychotherapy, Bern University, Bern, Switzerland

## Abstract

**Background:** A major challenge in neurodegenerative diseases concerns identifying biological disease signatures that track with disease progression or respond to an intervention. Several clinical trials in Huntington disease (HD), an inherited, progressive neurodegenerative disease, are currently ongoing. Therefore, we examine whether peripheral tissues can serve as a source of readily accessible biological signatures at the RNA and protein level in HD patients.

**Results:** We generate large, high-quality human datasets from skeletal muscle, skin and adipose tissue to probe molecular changes in human premanifest and early manifest HD patients—those most likely involved in clinical trials. The analysis of the transcriptomics and proteomics data shows robust, stage-dependent dysregulation. Gene ontology analysis confirms the involvement of inflammation and energy metabolism in peripheral HD pathogenesis. Furthermore, we observe changes in the homeostasis of extracellular vesicles, where we find consistent changes of genes and proteins involved in this process. In-depth single nucleotide polymorphism data across the HTT gene are derived from the generated primary cell lines.

**Conclusions:** Our omics data document the involvement of inflammation, energy metabolism, and extracellular vesicle homeostasis. This demonstrates the potential to identify biological signatures from peripheral tissues in HD suitable as biomarkers in clinical trials. The generated data, complemented by the primary cell lines established from peripheral tissues, and a large panel of iPSC lines that can serve as human models of HD are a valuable and unique resource to advance the current understanding of molecular mechanisms driving HD pathogenesis.

**Keywords:** Neurodegeneration, Transcriptomics, Proteomics, SNP sequencing, Skeletal muscle, Adipose tissue, Skin, iPSC



© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

## Background

The CAG repeat expansion mutation in exon 1 of the huntingtin gene (*HTT*) causes Huntington disease (HD), a progressive movement disorder with dementia and behavioral abnormalities [1]. Huntingtin is expressed ubiquitously throughout the body, and HD does not exclusively affect the brain. In addition to a particular vulnerability of striatal and cortical neurons to the mutation, several lines of evidence suggest that mutant HTT influences molecular function in peripheral tissues including skeletal muscle, peripheral blood monocytes, liver, and others [2]. Hence, the investigation of peripheral tissues and cells holds promise to reveal important insight into the physiological function of HTT that may include transcriptional processes, protein trafficking, and vesicle transport [3, 4]. It can also reveal peripheral signatures of HD that track with the evolution of the expression of HD in the central nervous system. This could have potential as an easily accessible and low invasive biomarker in clinical trials in HD. Finally, primary cell cultures established from peripheral tissues such as skin can be used to generate pluripotent stem cells which, when differentiated into neuronal and non-neuronal cells, can serve as human models of HD.

The objective of the Multiple Tissue Monitoring in Huntington disease (MTM-HD) study was to comprehensively examine the molecular biology at the RNA and protein level of peripheral tissues from *HTT* CAG repeat expansion mutation carriers and sex and age matched healthy volunteers. All participants were also participants in Enroll-HD ([www.enroll-hd.org](http://www.enroll-hd.org)), and extensive demographic and phenotypic data were collected according to the Enroll-HD protocol. The analyses of proteomics and transcriptomics datasets in conjunction with the available deep clinical phenotype information can give valuable insights into more complex signatures in human carriers of the HD gene mutation and provide a resource of peripheral tissues and primary cells with accompanying multi-layer 'omics datasets.

The main potential of these types of analyses is in relating biological signature(s)—either in individual modalities and/or tissues or across modalities and tissues—with CAG repeat length as the severity of the expression of the HD gene mutation. Lastly, the insight gained through these analyses can inform future research with the aim to better understand the nature of any peripheral biological signature and the impact an intervention, e.g., in a clinical trial, can have on them.

## Results

Skeletal muscle, adipose tissue and skin from an open biopsy of the *quadriceps femoris* muscle, as well as blood, was collected from 20 healthy controls, 21 pre-symptomatic, and 20 early motor manifest HD (UHDRS total functional capacity stages 1 and 2) patients. Additionally, primary fibroblast and myoblast cell lines have been established. In the following we present the demographic and clinical data for the participants, the initial findings of transcriptional and proteomic dysregulation in tissues, as well as the analysis of the generated fibroblasts by RNAseq and *HTT*-centered single nucleotide polymorphism (SNP) sequencing. The analysis of the epigenetic clock in the fibroblast lines has already been published [5]. As an additional resource,

pluripotent stem cell lines (iPSCs) have been generated from the primary fibroblast lines and are available for further studies.

#### Participant's demographic and clinical data

At both sites, a total of 24 unaffected individuals (control), 23 pre-manifest (known mutation carriers, before disease onset; pre-HD), and 21 early HD (early-HD) patients were recruited into the study that was approved by the Ethics Review Boards of Ulm University and University College London. Following a detailed description of the study participants gave informed consent according to the Declaration of Helsinki. Participants attended two visits: the screening and sampling visits, which were no more than 1 month apart. At the screening visit, participants were assessed for inclusion and exclusion criteria (Table 1). Ultimately, 20 control, 21 pre-HD, and 20 early-HD participants were included in the sampling visit of the study. Summary data for these 61 participants are shown in Table 2. For the full dataset including statistics, see Additional file 1 (MTM\_all\_metadata.xlsx). Participants from all three groups were evenly recruited across the two sites and were matched for gender, as well as age (Table 2). Neither age, weight, height, nor BMI were significantly different between the groups (one-way

**Table 1** Inclusion and exclusion criteria for participants of the MTM-HD study

Inclusion criteria	Exclusion criteria
Age 30–50	Major psychiatric disorder at time of enrolment
Caucasian ethnicity to minimize genetic differences in this relatively small sample set	On long-term anti-coagulants, e.g., Warfarin for mechanical heart valves, pulmonary embolism or atrial fibrillation
Huntington disease gene CAG repeat length 40–55, for gene expanded subjects only	On combination of aspirin and clopidogrel for ischemic heart disease
Premanifest subjects stratified as closer to predicted disease onset (disease burden score > 250, UHDRS diagnostic confidence score < 3)	Any relevant condition (including presence of psychosis and/or confusional states), behavior, laboratory value, or concomitant medication which, in the opinion of the investigator, makes the subject unsuitable for entry into the study
Early HD subjects (stage 1 and 2 disease as quantified by UHDRS total functional capacity score 7–13) aged 30–50	Recent major abdominal or limb surgery within last month
Healthy volunteers aged 30–50	Fear of needles
REGISTRY or ENROLL-HD assessment within 6 months of screening visit, for gene expanded subjects only	Anemia (Hb < 11.5), thrombocytopenia (platelets < 150), or coagulopathy (INR > 1.5) identified on screening and/or blood tests
Presence of companion to accompany them home	High-dose nutraceuticals, e.g., creatinine, co-enzyme Q, vitamin E
Capable of providing informed consent	Significant lower limb impairment of any kind
Subjects must have no clinically significant and relevant history that could affect the conduct of the study and evaluation of the data, as ascertained by the investigator through detailed medical history	Significant cardio-respiratory or other medical comorbidities such as ischemic heart disease, diabetes, chronic obstructive pulmonary disease, chronic liver disease, stroke with hemiparesis, BMI > 30, and history of malignancy in last 2 years
Ability to tolerate procedures, fasting and blood sample donation	Illicit drug use and/or alcohol abuse > 21 units/week males; 14 units/week female
	Participation in an investigational drug trial within 30 days of screening visit
	Pregnancy
	Allergy/previous adverse reaction to local anesthetic agents

**Table 2** Summary data for the 61 participants of the MTM+HD study

HD status	Site	Gender	Age [years]	Short CAG allele	Long CAG allele	DBS	Weight [kg]	Height [cm]	BMI	TFC	TMS	FA	IS
Control	1: 10	f: 10	41.3	15.2	18.5	n.a.	74.4	171.7	25.1	13	0.5	25	100
	2: 10	m: 10	[31.6; 55.9]	[10;20]	[15;27]		[45.3;102]	[158;187]	[17.6;33.7]	[13;13]	[0;4]	[25;25]	[100;100]
Pre-HD	1: 10	f: 10	41.6	19.6	43.6	327.7	76.2	172.8	25.4	12.7	5.3	24.7	97.6
	2: 11	m: 11	[31.2; 56.5]	[9;24]	[40;48]	[158.6; 423.8]	[49.2;115]	[154;190]	[20.4;35.5]	[11;13]	[0;18]	[22;25]	[80;100]
Early-HD	1: 10	f: 10	45.4	18.9	44.5	406.7	73.7	173.5	24.4	10.9	24	22.3	88.4
	2: 10	m: 10	[35.4; 58.9]	[9;25]	[41;50]	[281.0; 657.3]	[55;94.4]	[153;197]	[18.6;27.9]	[7;13]	[8;49]	[16;25]	[70;100]

Data are mean and minimal and maximal values, respectively, in brackets. *Abbreviations:* Gender: f female, m male. *DBS* disease burden score (not applicable (n.a.) in the control group). *BMI* body mass index. *TFC* total functional capacity. *TMS* total motor score, *FA* function analysis, *IS* independence scale



both tissues. In both tissues, the gene was upregulated in the early-HD group compared with the pre-HD group (adipose: log<sub>2</sub>-fold change 4.47,  $p_{\text{adj}} = 0.003$ ; muscle: log<sub>2</sub>-fold change 1.51,  $p_{\text{adj}} = 0.041$ ). In both tissues, sets of genes were exclusively dysregulated in either the pre-HD or early-HD groups (Fig. 1A, B, pre-HD only and early-HD only).

Gene ontology enrichment in adipose tissue for the different comparisons showed that REST (RE1-silencing transcription factor) target genes are dysregulated only in the pre-HD group (Fig. 1A), while protein sumoylation was only dysregulated in the early-HD group (Fig. 1A). Alteration in protein sumoylation is a well described mechanism in HD, both for HTT itself, as well as in general [6–11]. Also, REST has been implicated in HD pathogenesis [12–14].

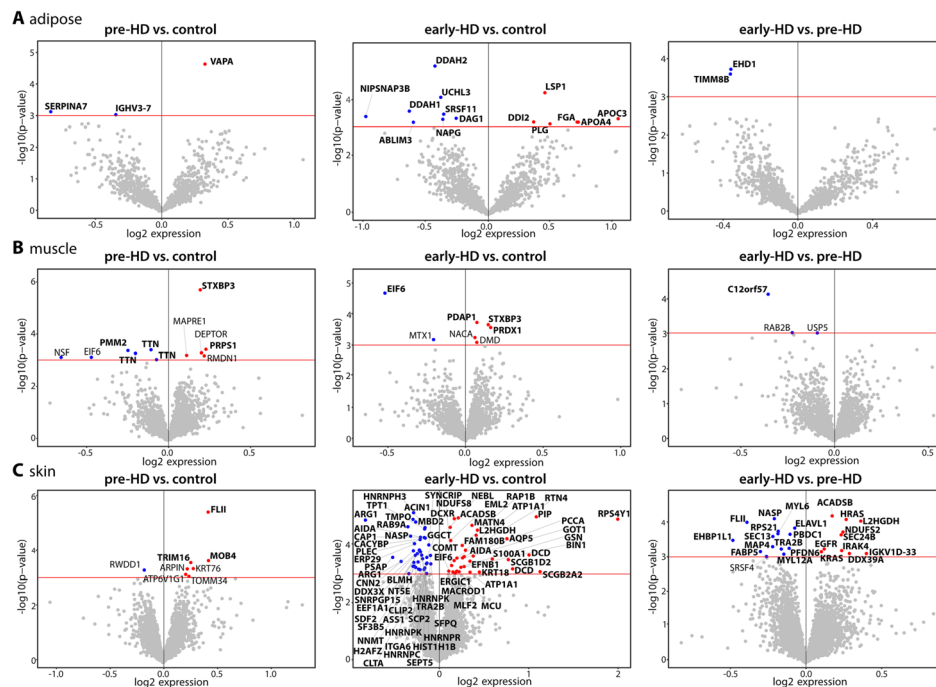
Gene ontology enrichment analysis in muscle points towards disrupted homeostatic pathways, especially in the pre-HD group (Fig. 1B). Interestingly, *PAX6*, an important regulator of diverse peripheral and central nervous system processes, was highly and progressively upregulated in both HD groups (Fig. 1B, similar regulation), suggesting a compensatory mechanism of muscle regeneration in response to mutant HTT expression. Similar to our finding, in the muscle of R6/2 mice [15], a higher level of satellite cells (skeletal muscle progenitor cells crucial for repair and regeneration) has been found, concomitant with an increase in *PAX7*, another member of the PAX transcription factor family that is key for normal satellite cell function in skeletal muscle regeneration [16]. The enrichment analysis of the dysregulated genes in the muscle pre-HD group suggested *PPARA* (peroxisome proliferator activated receptor alpha) as a regulatory protein with very high confidence (Fig. 1B).

#### Transcriptomic analysis of fibroblast lines

We generated 100 bp long, paired-end reads with a depth of approximately 40 to 60 million reads per sample of polyA-enriched total RNA (RNAseq counts: Additional file 6; DESeq2 analysis Additional file 7). Evaluation was performed in the same way as described for tissues above. Only few genes were significantly dysregulated between pre-HD and early-HD groups as compared to controls (Fig. 1C). Consequently, gene ontology analysis resulted in few enriched terms (Fig. 1C). Notably, *TBC1D3D* (ENSG00000274419), which was found to be dysregulated in both the adipose and muscle datasets (see above and Fig. 1A, B), was also dysregulated in the fibroblast dataset. Its expression pattern was consistent between all three datasets with a downregulation in the pre-HD group and a restoration towards control levels in the early-HD group.

#### Proteomic analysis of skin, adipose, and muscle tissue

Next, we generated proteomics data from the skin, adipose tissue and muscle biopsies. In adipose tissue, we identified 1347 proteins (Additional files 8 and 9), in muscle 2671 proteins (Additional files 10 and 11), and in skin 4640 proteins (Additional files 12 and 13). After correction for confounding variables (TMT-plex, site of sampling, gender, age and BMI; see the “Methods” section for details), we performed dysregulation analysis using ROTS (reproducibility-optimized test statistic) [17]. ROTS uses a non-predetermined optimized test statistic for each dataset by utilizing bootstrapped datasets that preserve the individual top-ranked features and maximizing the overlap between those. The data files for peptide and protein analysis can be found in the Additional files 14 and



**Fig. 2** Proteomics analysis of adipose, muscle, and skin samples. **A–C** Volcano plots of proteomics data from adipose (**A**), muscle (**B**), and skin (**C**) tissue samples. Pair-wise comparisons of pre-HD and early-HD against controls and early-HD against pre-HD are shown. The x-axis is the  $\log_2(\text{expression})$  change and y-axis  $-\log_{10}(p\text{-value})$  from the ROTS analysis. The red horizontal line highlights the  $p$ -value of  $-\log_{10}(3) = 0.001$ . Proteins with a significance value of less than 0.001 are depicted and proteins with an adjusted  $p$ -value (ROTS FDR) of  $< 0.05$  are shown in bold. Upregulated protein are shown in red, downregulated proteins in blue. See also Additional files 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18 and 19 for full data

15 (adipose), Additional files 16 and 17 (muscle), and Additional files 18 and 19 (skin). The analyses are shown as volcano plots for each of the pairwise comparisons for each tissue in Fig. 2 with significantly dysregulated proteins ( $p$ -value  $< 0.001$ ) highlighted. Proteins that were also determined to be dysregulated based on analysis of permuted sample lists are highlighted in bold (FDR  $< 0.05$ ).

In summary, for all three tissues we observed a progressive increase of the number of dysregulated proteins from pre-HD to early-HD stage. In adipose tissue, there was no overlap of commonly dysregulated proteins in the three comparisons (Fig. 2A). In muscle, syntaxin-binding protein 3 (STXBP3) was upregulated in both pre- and early-HD samples (Fig. 2B). We observed the largest changes in skin (Fig. 2C). Here, MOB family member 4, phocein (MOB4) was upregulated in both, pre- and early-HD samples. One hundred sixty-two proteins were significantly (FDR  $< 0.05$ ) dysregulated in the early-HD group. Twenty-three proteins significantly changed their expression levels from pre-HD to early-HD stage (Fig. 2C).

Due to the low number of changed proteins, we only performed gene ontology enrichment in the adipose tissue dataset (early-HD compared to controls) and the skin dataset (early-HD compared to controls and the comparison of pre-HD to early-HD) (Table 3). Pathway and ontology analysis in the adipose tissue dataset hinted at dysregulated lipid metabolism and proteasome function (Table 3, adipose: early-HD vs. controls). Both



**Table 3** Ontology enrichment of proteomics data

Dataset	Pathways	Ontology [biological process]	Regulators
	1: WikiPathways 2019 (human) 2: KEGG 2019 (human)		1: ChEA 2016 2: TTRUST 2019 3: ENCODE 2015
Adipose: early-HD vs. control	1: Blood clotting cascade (1480.08), statin pathway (408.90), proteasome degradation (378.24) 2: Complement and coagulation cascades (546.42), proteasome (225.26), cholesterol metabolism (195.08)	Chylomicron remodeling (2106.52), nitric oxide-mediated signal transduction (1008.13), fibrinolysis (921.53), proteasomal ubiquitin-independent protein catabolic process (528.58), interleukin-1-mediated signaling pathway (209.32)	1: HNF4A (14.55), AR (13.99), NANOG (13.59) 2: PPARA (273.66) 3: FOXA2 (29.83), NR2C2 (25.90), HNF4G (24.99)
Skin: early-HD vs. control	1: Translation factors (337.17), mRNA processing (261.60), methylation pathways (234.97) 2: Spliceosome (298.11), arginine biosynthesis (167.83), tyrosine metabolism (72.16)	Protein localization to cytoplasmic stress granule (356.98), mRNA splicing, via spliceosome (343.79), IRES-dependent viral translational initiation (198.37), aspartate family amino acid metabolic process (198.37), nucleoside metabolic process (198.37)	1: XRN2 (23.62), AR (23.21), TP63 (22.90) 2: BRCA1 (34.50), VDR (21.17), AR (14.95) 3: NELFE (93.56), KAT2A (74.77), TAF7 (48.19)
Skin: early-HD vs. pre-HD	1: Extracellular vesicle-mediated signaling in recipient cells (1346.70), ErbB signaling pathway (298.26), AGE/RAGE pathway (176.44) 2: ErbB signaling pathway (327.55), oxytocin signaling pathway (295.74), neurotrophin signaling pathway (205.61)	Cargo loading into COPII-coated vesicle (1787.10), ERBB2 signaling pathway (942.71), ephrin receptor signaling pathway (130.95), Fc receptor signaling pathway (111.42), mRNA splicing, via spliceosome (65.83)	1: SOX11 (22.22), EGR1 (22.10), FOXP2 (19.42) 2: TP53 (130.54), SP1 (9.04) 3: TEAD4 (60.02), GTF3C2 (23.76), SP2 (20.43)

Ontology enrichment using the Enrichr website. See the “Methods” section for details. Only non-redundant, aggregated terms are shown. Only entries of human samples in the databases were considered. Only terms that applied to at least two dysregulated proteins were considered. The combined score is shown in brackets. Only entries with a combined score of 5 or greater were considered. Top three terms for pathways and regulators and top 5 terms for ontology are shown

have been extensively studied in HD. Peroxisome proliferator activated receptor alpha (PPARA) was strongly predicted as potential upstream regulator of these dysregulated proteins (Table 3). Worth noting, PPARA was also predicted as the most likely upstream regulator for the dysregulated gene analysis of the muscle dataset (Fig. 1B).

Strikingly, 21 out of the 28 dysregulated proteins in the adipose tissue dataset (early-HD vs. controls) are associated with the term “extracellular vesicle.” Extracellular vesicles (EVs) are composed of lipids, proteins, and RNAs, are membrane coated, and are secreted by most cells. Moreover, they are implicated in neurodegeneration, although the availability of data for HD is poor [18].

Pathway and ontology enrichment in the skin proteomics dataset showed that proteins related to gene expression in general are affected in the early-HD samples compared to controls (Table 3). RNA processing and translation as well as amino acid metabolism seems to be dysregulated. The predicted regulators are all factors involved in gene expression. Of note, androgen receptor (AR) and lysine acetyltransferase 2A (KAT2A), the latter in a complex with ataxin 7 (ATXN7), are both linked to polyglutamine diseases. A CAG expansion in AR causes spinal and bulbar muscular atrophy [19]; an expansion in ATXN7 causes spinocerebellar ataxia type 7 [20]. This could potentially indicate that



there are common phenomena caused by a CAG repeat expansion, independent of the mutated gene. The enriched pathways in the comparison of the early-HD and pre-HD samples in the skin dataset (Table 3) are intertwined and regulate cell survival and proliferation. Consequently, TP53 (tumor protein P53), a key regulator of cellular survival, was predicted as an upstream regulatory protein. In addition, extracellular vesicle mediated signaling was again predicted with a huge combined score (Table 3).

### Other resources generated under the MTM-HD study

#### *HTT gene SNP sequencing*

An important consideration in therapeutic *HTT* lowering trials is preservation of a certain level of HTT protein. While inactivation of HTT in the brain of adult mice [21] and primates [22] seems to be tolerated for some time, inactivation of HTT beyond a certain level, likely 50% of the normal level, in the developing, as well as the adult brain potentially has severe consequences (reviewed in [23, 24]). To avoid indiscriminately reducing both wild-type and mutant HTT, an attractive approach in human *HTT* lowering trials is to target the alleles of single nucleotide polymorphisms (SNPs) that are phased on the mutated *HTT* allele. To facilitate such approaches, we genome-sequenced the primary fibroblast lines generated in this study by the 10X Genomics Chromium technology. We only sequenced the lines that were generated from HD patients ( $n = 39$ ). For patient privacy and legal reasons, here we only show the *HTT* gene ( $\pm 100$  kbp) associated SNPs (Additional file 20). Overall, 98.3% of all identified SNPs could be assigned to a specific haplotype. The largest block of assembled *HTT* sequencing data was around 14.5 million base pairs.

We previously defined 16 haplotypes that are common in HD subjects with European ancestry [25] and demonstrated allele-specific target sites for each haplotype [26, 27]. To maximize the discrimination power of the haplotypes and to cover non-Europeans, we extended the *HTT* haplotype definitions by using 1000 Genomes Project (KGP) data (phase 3, all populations). Based on the same 21 genetic variations that were used to define the original 16 haplotypes [25], 253 additional haplotypes were identified and subsequently assigned based on the allele frequency in all KGP populations. For example, excluding the original 16 haplotypes, hap.17 represents the most frequent *HTT* haplotype in KGP data, accounting for approximately 3% of haplotypes in KGP chromosomes. The list of all newly defined HD haplotypes can be found in the Additional file 21.

Out of the total of 39 sequenced samples, the *HTT* haplotypes of 37 samples could be conclusively defined (Additional Fig. S1A and B). A phylogenetic tree analysis on the *HTT* haplotypes to study evolutionary relationships between the different *HTT* alleles revealed three major groups of *HTT* haplotypes (represented by colors in Additional Fig. S1C and D).

#### *Genome-wide methylation data*

DNA methylation states, which are part of the epigenetic modification landscape, allow the assessment of the “molecular age” of a cell through the analysis of epigenetic clocks [28]. The fibroblast lines described in this manuscript were part of a multi-species study

to analyze changes in the methylome in the background of HD [5]. Associated data has been published by the authors [5].

### **Generation of iPSC lines**

In addition to the 'omics datasets as mentioned above, we also generated, as a flexible model for HD research, a total of 45 iPSC lines. These lines were derived from the primary fibroblast cell lines in this study (controls and HD patients). All fibroblast lines were used to generate iPSC lines. However, for technical reasons, only 45 lines resulted in viable iPSC lines that also passed quality control. The iPSC lines are available from the European Bank for Induced pluripotent Stem Cells (EBiSC; search for "CHDI" at <https://cells.ebisc.org>). Re-programming of fibroblasts was conducted using the non-integrative Sendai virus strategy expressing KLF4, MYC, POU5F1 (OCT4), and SOX2 [29]. All lines were characterized regarding expression of pluripotency markers (OCT-4 (POU5F1), TRA-1-60, SSEA-1, and SSEA-4), retention of re-programming vector sequences, morphology, and karyotype. Furthermore, iPSC lines were differentiated into endo-, meso-, and ectodermal cells and marker expression was analyzed (endoderm: CXCR4, GATA6, SOX17; mesoderm: MIXL1, NCAM1, VIMENTIN; ectoderm: HES5, NEUROD1, PAX6). All data are freely available from the EBiSC website (at the time of May 2022). The *HTT* haplotype information of the iPSC lines is linked to the EBiSC identifier in the Additional file 20.

### **Discussion**

Many valuable insights into the pathogenesis of HD have been gained from disease models, whether in mice, flies, or non-human primates [30, 31] to suggest that the expression of the HD mutation affects the innate immune system and energy metabolism [32]. The in-depth molecular tissue signatures, which we derived from human HD patients' peripheral tissues specifically indicate PPARA (peroxisome proliferator activated receptor alpha) as a regulatory protein with very high confidence for dysregulated genes in muscle in the pre-HD group (Fig. 1B) and for dysregulated proteins in adipose tissue in the early-HD group (Table 3). In other, independent, external datasets PPARA was found to be dysregulated itself: It was significantly upregulated in post mortem caudate and cerebellum [33] and also significantly upregulated in post mortem pre-frontal cortex [34]. While none of our datasets indicate that PPARA itself is statistically significantly dysregulated in muscle, adipose, or skin, very small changes in regulatory proteins can have a much larger knock-on effect on target genes, as suggested by our analysis. The family of PPA receptors are important regulators of various processes including inflammation, lipid and energy metabolism. The PPA receptors interact with, e.g., PGC-1 $\alpha$  (peroxisome proliferator-activated receptor gamma coactivator) to enable the correct function of both partners including in the central nervous system [35]. Paralogs of PPARA, namely PPARD [36, 37] and PPARG [38, 39], have been implicated in HD pathogenesis. Moreover, PPARs can be therapeutically targeted to improve phenotypes such as mitochondrial dysfunction in HD model systems [37–40].

The most consistently changed gene in the RNAseq data from muscle, adipose, and fibroblasts was *TBC1D3D* (ENSG00000274419). *TBC1D3D* is not included in published

'omics datasets (see above); however, in all tissues, *TBC1D3D* expression was significantly downregulated in the pre-HD group and restored towards control levels in the early-HD group in our datasets (Fig. 1). *TBC1D3D* protein is expressed in all three tissue types ([www.uniprot.org](http://www.uniprot.org), [41]) and most probably acts as a GTPase activating protein to stimulate RAB5 function [42]. RAB5 is an early endosome marker and functions in endocytosis and membrane transport where it regulates receptor sorting and vesicle fusion [43]. Intriguingly, the HTT-HAP40 complex also acts as a regulatory protein complex for RAB5 [44]. Consequently, expression of mutant HTT leads to aberrant endosomal/lysosomal pathway function [45] and its secretion via a non-canonical pathway [46]. Endosomal pathway integrity is a prerequisite for functioning secretion of exosomes through multi-vesicular bodies (part of the endosomal system) and vice versa [47]. Exosomes are one of the three main species of extracellular vesicles (EVs), the others being microvesicles, also called ectosomes, and apoptotic bodies [48]. Apoptotic bodies form during cell death; microvesicles directly bud from the cell plasma membrane. All EVs are composed of lipids, proteins, RNAs, and other small molecules. Their circulating nature makes them attractive messengers of information from the secreting cell towards the target cell. They are also able to cross the blood brain barrier and are found in virtually every bodily fluid [48]. When we analyzed the dysregulated proteins in the adipose dataset (early-HD vs. controls), we found an exceptionally high association of the gene ontology term “extracellular vesicle” (for 21 out of the 28 dysregulated proteins) (Table 3). Additionally, gene ontology enrichment of dysregulated proteins in the skin dataset also pointed towards “Extracellular vesicle-mediated signaling in recipient cells” (Table 3). Extracellular vesicles could play an important role in neurodegeneration as they have been implicated as one route for the spread of misfolded proteins, for instance in Parkinson’s disease, and they also contain other cargo such as miRNAs [49]. In HD, our human tissue data indicate that the expression of the HD mutation affects the homeostasis of extracellular vesicles similar to what has been observed before [50]. These findings certainly merit further investigation, in particular because EVs derived from peripheral tissues and biofluids are an attractive source of (peripheral) biomarkers for clinical trials.

## Conclusions

In summary, the biological signatures of the changes at RNA and protein levels point towards the involvement of inflammation, energy metabolism, and vesicle biology in peripheral tissues in HD. It shows the potential to identify biological signatures from peripheral tissues in HD that could be suitable as biomarkers in clinical trials. In addition, we have generated a high-quality human multi-omics reference data set and established valuable primary cell lines with associated *HTT* haplotyping data and generated iPSC cell lines. All participants took part in a large longitudinal observational study, Enroll-HD, with deep phenotyping [51]. These data provide an important resource for further research into, for instance, somatic repeat instability, cell type-specific effects of mutant HTT, or the HD epigenome [5]. SNP-based HTT lowering approaches are underway to selectively target only the expanded allele harboring the SNP [23, 24, 52]. Our fully SNP phased fibroblasts and the iPSC lines derived from them can be used to analyze the feasibility of these approaches and to develop novel molecules for therapy.

iPSC lines are also valuable tools in many neurodegenerative disorders [53], including HD [54], and our set of well-characterized lines can be used to study pathomechanisms across diseases.

## Methods

### Study participants

Participants of the MTM-HD study were recruited at the departments of neurology of Ulm University (Germany) and University College London (United Kingdom). At both institutions the study followed a standardized protocol. Participants were only included if there were no contra-indications to muscle biopsy, e.g., a clotting disorder or evidence to suggest cardio-respiratory abnormalities. The only relatives in the cohort were a pair of sisters in the Ulm samples. None of the other individuals were related. We therefore did not consider the degree of relationship in our analyses. For HD patients, the CAG-tract length was determined and participants were clinically assessed as described for the TrackOn and TRACK-HD studies [55, 56]. Clinical assessment included the United Huntington Disease Rating Scale (UHDRS) motor part to derive the total motor score and the UHDRS total functional capacity scale (TFC) [57]. The disease burden score (DBS) was calculated from each HD participant's CAG repeat length and age according to the following formula:  $(\text{CAG}-35.5) \times \text{age}$  [58]. A DBS of 250 or greater was used to screen for potential HD participants. Furthermore, HD participants were categorized as pre-HD if they had a diagnostic confidence level score of 2 or less on the UHDRS motor scale, or as early-HD if they were in TFC stages of 1 or 2, indicative of early motor manifest HD. Only participants with no signs of neuromuscular disease on clinical examination were enrolled. The local ethics committees at Ulm University and University College London approved the study (Ulm: 265-12; London: 12/LO/1565), and written informed consent was obtained from each participant. Following informed consent, participants were given a 9-digit unique identifier as described [59].

### Tissue biopsy and sample collection

In order to ensure samples are collected and processed in exactly the same manner in Ulm and in London, personnel at either site received on-site training on all aspects of the protocol. All samples were collected following an overnight fast (water was permitted) between 7.30 and 9 am local time (CET or GMT) to account for any influence of circadian rhythm. Open biopsies of the skin, subcutaneous adipose tissue and *M. vastus lateralis* were obtained following local anesthesia. Participants were contacted on the day after and 7 days following the procedures. After 1 week, the stitches were removed. Tissues were dissected (skeletal muscle, adipose tissue and skin), immediately snap-frozen in liquid nitrogen and stored at  $-80^{\circ}\text{C}$ . Blood samples were taken following informed consent on the screening day. This included routine blood tests (full blood count, renal, liver, bone and clotting profiles, creatinine kinase and group and save). In addition, plasma and buffy coat were separated. PBMCs were extracted from CPT/heparin tubes (BD Vacutainer CPT) as previously described [60]. Cell pellets were then snap frozen in liquid nitrogen and stored at  $-80^{\circ}\text{C}$ . For histochemistry and immunohistochemistry, the muscle sample was mounted on a piece of cork in TissueTek with

fibers oriented perpendicular to the cork and then snap-frozen in liquid nitrogen-cooled 2-methylbutane and stored at  $-80^{\circ}\text{C}$ . For electron microscopic analysis, a muscle fiber was pinned onto a cork plate with two needles in the operating theatre and immediately fixed in 2.5% glutaraldehyde.

#### **Generation of primary cell lines**

Primary human fibroblast and myoblast cell lines from the MTM participants were established and maintained as previously described [61].

#### **RNA and DNA extraction**

RNA extraction from tissues was performed by Labcorp (NC, US) using the Qiagen miRNeasy kit. RNA had to pass the following specifications from an Agilent Bioanalyzer run to be used for subsequent sequencing: 28S/18S ratio: 0.8–3.0 and RIN score:  $\geq 6.0$ . For fibroblast lines, RNA and DNA were extracted simultaneously by Q Squared Solutions Expression Analysis LLC (NC, US). The Qiagen AllPrep DNA/RNA kit was used to simultaneously purify genomic DNA and total RNA from the same sample. Briefly, samples were lysed in a guanidine-isothiocyanate containing buffer and the lysate was passed through an AllPrep DNA spin column which selectively binds genomic DNA. The column was washed and the genomic DNA was eluted in EB buffer stored at  $-20^{\circ}\text{C}$ . Ethanol was added to the flow-through of the AllPrep DNA spin column which allows for efficient binding of RNA. The mixture was then applied to a RNeasy spin column (Qiagen), where total RNA binds to the membrane. RNA was then eluted in RNase-free water, quantified and integrity assessed using the RNA 6000 Nano Assay on a Bioanalyzer 2100. RNA passed QC with RIN values  $\geq 7.0$ . The purified RNA was stored at  $-80^{\circ}\text{C}$ .

#### **RNA sequencing of adipose and muscle tissue RNA**

RNA sequencing was performed by Q Squared Solutions Expression Analysis LLC (NC, US). RNA samples were converted into cDNA libraries using the Illumina TruSeq Stranded Total RNA sample preparation kit (Illumina # RS-122-2303). Briefly, total RNA samples were concentration normalized, and ribosomal RNA (rRNA) was removed using biotinylated probes that selectively bind rRNA species. The resulting rRNA-depleted RNA was fragmented using heat in the presence of divalent cations. Fragmented RNA was converted into double-stranded cDNA, with dUTP utilized in place of dTTP in the second strand master mix. A single “A” base was added to the cDNA and forked adaptors that include index, or barcode sequences were attached via ligation. The resulting molecules were amplified via polymerase chain reaction (PCR). During PCR, the polymerase stalls when a dUTP base is encountered in the template. Since only the second strand includes the dUTP base, this renders the first strand the only viable template, thereby preserving the strand information. Final libraries were quantified, normalized, and pooled. Pooled libraries were bound to the surface of a flow cell and each bound template molecule was clonally amplified up to 1000-fold to create individual clusters. Final libraries were sequenced on the

HiSeq 2500 Illumina sequencing platform with paired-end, 50 bp long reads with a total read depth of approximately 150M reads per sample. Sequencing raw files were demultiplexed, adapter trimmed, and clipped for low quality base calls. Final fastQ files were produced with reads of a minimum length after clipping/trimming of 25 nucleotides.

#### **RNA sequencing of fibroblast RNA**

RNA sequencing was performed by Q Squared Solutions Expression Analysis LLC (NC, US). Sequencing libraries were created using the Illumina TruSeq Stranded mRNA method (Illumina, RS-122-2103), which preferentially selects for messenger RNA by taking advantage of the polyadenylated tail. In summary, approximately 100 ng of total RNA per sample was used to purify poly-adenylated RNAs using oligo-dT attached to magnetic beads. Purified mRNAs were fragmented using heat in the presence of divalent cations. The fragmented RNAs were converted into double-stranded cDNA, with dUTP utilized in place of dTTP in the second strand master mix. This facilitates the preservation of strand information, as amplification in the final PCR step will stall when it encounters uracil in the nucleotide strand, rendering the first strand as the only viable amplification template. The double-stranded cDNA underwent end-repair, A-tailing, and ligation of adapters that include index sequences. The final libraries were amplified via polymerase chain reaction (PCR), after which they were quantified, normalized, and pooled in preparation for sequencing. Normalized libraries were multiplexed for efficient sequencing to the required number of reads per sample. Pooled libraries were bound to the surface of a flow cell, and each bound template molecule was clonally amplified up to 1000-fold to create individual clusters. Libraries were sequenced using the Illumina sequencing-by-synthesis platform, with a sequencing protocol of 100 bp paired-end sequencing and total read depth of 40M reads per sample on a NovaSeq. Sequencing raw files were demultiplexed, adapter trimmed, and clipped for low quality base calls. Final fastQ files were produced with reads of a minimum length after clipping/trimming of 25 nucleotides.

#### **Bioinformatics analysis of RNAseq datasets**

FastQ files were quality checked with FastQC v0.11.5 [62]. All files passed QC. The reads were aligned against Ensembl *homo sapiens* GRCh38 release 90 using STAR aligner v2.5.3a [63]. Reads were quantified using salmon v0.8.2 [64]. Samples were screened for outliers using a combined PCA and clustering analysis. A sample was defined as an outlier if it was outside a 68% probability ellipse in PCA analysis and was outside a  $-2.5$  standardized connectivity cutoff of a Euclidean distance matrix of all samples per each tissue [65]. This procedure identified 6 samples in the adipose ( $n = 2$  for each control, pre-HD and early-HD), 3 samples in the muscle ( $n = 1$  control,  $n = 2$  pre-HD), and 10 samples in the fibroblast ( $n = 4$  control,  $n = 4$  pre-HD,  $n = 2$  early-HD) datasets. For all subsequent analyses, both RNAseq batches for muscle and adipose tissues were combined. Transcriptional dysregulation was computed using tximport v1.10.0 [66] and DESeq2 v1.22.1 [67]. The DESeq2 intercept design normalized counts are available as supplementary information files. For the dysregulation analysis, HD clinical stage



(control, pre-HD, early-HD) was used as the variable of interest. RNAseq batch (for tissues), gender, site of sampling, age, and BMI were used as covariates in the modeling. We factorized the two continuous variables (age and BMI) into 5 intervals before evaluation with the cut function in R. Ashr was used as the fold change shrinkage estimator [68]. DESeq2 analysis files are available as supplementary information files. Ontology analysis was carried out using the Enrichr website [69, 70].

### **Proteomic analysis and bioinformatics**

Proteomics analysis was performed by proteome sciences ([www.proteomics.com](http://www.proteomics.com)). For TMT labeling (Thermo Fisher), 2 mg of muscle tissue were reduced (dithiothreitol), alkylated (iodoacetamide), digested (trypsin) to generate peptides, desalted (SepPak tC18 cartridges (Waters, Milford, MA, USA)), and lyophilized. Peptides were mixed with their respective TMT 10plex tag and incubated for 1 h at room temperature. Individual TMT reactions were terminated with hydroxylamine and the labeled digests of each of the 10 samples were pooled into the respective TMT 10plex and incubated for another hour. Finally, the TMT 10plex analytical sample was acidified, diluted to an acetonitrile concentration less than 5%, divided into two aliquots, desalted, and lyophilized to completion. Each TMT 10plex analytical sample aliquot was separated into 12 x 4-min fractions by strong cation exchange chromatography (polySULFOETHYL-A column (PolyLC) and HPLC system (Waters Alliance 2695)) and the 12 fractions combined by smart pooling into six fractions with roughly equal peptide amounts for subsequent processing. The muscle samples were subjected to two SCX-runs as the total load of the muscle TMT 10plexes was 20 mg each and makes two runs necessary. Re-suspended peptides were loaded onto a nanoViper C18 Acclaim PepMap 100 pre-column (Thermo Scientific). Each fraction was analyzed in duplicate by LC-MS/MS using the EASY-nLC 1000 system coupled to an Orbitrap FusionTM TribridTM Mass Spectrometer (both Thermo Scientific) for each of the TMT 10plex sets. Peptide mass spectra were acquired throughout the entire chromatographic run (180 min). MS-runs across the TMT 10plexes were submitted to Proteome Discoverer (PD) v1.4 (Thermo Scientific) using the Spectrum Files node. Spectrum selector was set to its default values, while the SEQUEST HT node was suitably set up to search data against the human FASTA UniProtKB/Swiss-Prot database (version 9606\_Human\_01082016 (created August 2016)). The reporter ions quantifier node was set up to measure the raw intensity values for TMT 10plex mono-isotopic ions (126, 127N, 127C, 128N, 128C, 129N, 129C, 130N, 130C, 131). The SEQUEST HT search engine was programmed to search for tryptic peptides (with up to two missed cleavages allowed) and with static modifications of carbamidomethyl (C), TMT6plex (K), and TMT6plex (N-Term). Dynamic modifications were set to deamidation (N/Q), oxidation (M), and phosphorylation (STY). Precursor mass tolerance was set to 20 ppm and fragment (b and y ions) mass tolerance to 0.02Da. The first stage of data processing was performed using the SQuaT Bioinformatics Module 3.1.1. Each TMT 10plex set was filtered to include PSMs (peptide spectral matches) that had a signal in at least one out of 10 TMT reporter ion channels. A correction procedure was then implemented to address isotopic impurities. Isotope correction factors used in this procedure were specific to the TMT batch used for labeling. Next, for every mass spectrometry run, the TMT reporter ion intensities were normalized using the sum-scaling technique at the



PSM level. Briefly, for each TMT channel all reporter ion intensities were summed and the median value across all summed channels was calculated. A correction factor was obtained by dividing the individual summed reporter ion intensity for a channel by the median of all channels. Finally, each individual PSM ion intensity was multiplied by the correction factor for the relevant TMT reporter channel. Post sum-scaling, the ratios of each sample PSM was related to the reference sample channel. The PSM level ratios to the reference sample were calculated as a log<sub>2</sub> transformation of sample PSM intensity relative to the reference sample PSM intensity. The data quality of the samples was controlled using quality control charts. Two quality control metrics per sample were calculated: the median (measure of central tendency) and the inter-quartile range (IQR) (measure of scale) using peptide and protein distributions. Both metrics are robust and stable to outliers and are therefore good for detecting extreme outliers. The median log expression and inter-quartile range (IQR), were traced using control charts for all datasets on peptide and protein level. A sample was considered as a strong outlier if either QC metric value was more than three standard deviations from the overall mean. A second stage of data processing was performed using the FeaST Bioinformatics Module 1.3.1. Imputation was performed to obtain quantitative values for the remaining missing data points using an iterative PCA method [71]. Following imputation, the data were normalized across the sets using quantile normalization method to center the sample values across the different TMT 10plex sets. Multifactorial limma [72] modeling was used to remove TMT 10plex set and site of sampling as batch effects. The final datafiles for adipose, muscle, and skin proteomics analysis are available as supplementary information files.

#### **Bioinformatics analysis of proteomics datasets**

For further analysis of the peptide datasets, all non-unique entries for peptides (shared\_status\_Gene) were removed, followed by removal of rows with unassigned protein IDs. We then used “Peptide” as the identifier to collapse rows (WGCNA function [73]) onto unique peptides entries. We next employed the same strategy as described above to identify outliers in the peptide and protein datasets but could not detect any outlying samples. Following this, we consecutively removed batch effects with limma [72] for both peptide and aggregated protein datasets. Adipose: gender, BMI, age; muscle: BMI, age; skin: age, gender, BMI. To compute peptide and protein dysregulation, we used ROTS [17]. For each pairwise comparison, we used the respective full dataset for modeling and ran 1000 bootstraps. Reproducibility Z-score value was greater than 2 for all comparisons. The ROTS analysis files are available as supplementary information files.

#### **SNP sequencing**

Genomic DNA extraction, sample QC, library preparation, sequencing reactions, and bioinformatics analyses were conducted at GENEWIZ, Inc. (South Plainfield, NJ, USA). High molecular weight genomic DNA was extracted from 2–4 million cells per sample, using Qiagen Genomic-tip 100/G HMW Kit (Qiagen, Hilden, Germany), according to manufacturer’s protocol. Sample amount was quantified using a Qubit 2.0 Fluorometer (Invitrogen, Carlsbad, CA, USA), and sample purity and integrity was checked using a Nanodrop spectrophotometer and a pulsed field gel analysis (PFGE) or an Agilent

TapeStation (Palo Alto, CA, USA), respectively. All samples were processed through the Chromium Controller following the standard manufacturer's specifications. DNA phasing libraries were generated from 1.07–2.41 ng of DNA per sample, using the 10X Genomics Microfluidic Genome Chip and 10X Genomics Chromium Genome kit (10X Genomics, CA, USA), according to manufacturer's protocol. The sequencing libraries were evaluated for quality using an Agilent TapeStation (Palo Alto, CA, USA), and quantified by using a Qubit 2.0 Fluorometer (Invitrogen, Carlsbad, CA, USA), and qPCR (Applied Biosystems, Carlsbad, CA, USA). The pooled sequencing libraries were loaded onto an Illumina HiSeq 4000 sequencer to achieve 150 Gb of data per sample. The samples were sequenced in a configuration compatible with the recommended guidelines as outlined by 10X Genomics (2 × 150 bp configuration, with 8 bp single index). A single sample with insufficient phasing results was sequenced to an additional 300 Gb of data. Raw sequence data (bcl) files generated by the machine were converted into fastq files and de-multiplexed using the 10X Genomics Cell Ranger software. Genome sequence variant calling, phasing and structural variant calling was done using 10X Genomics Long Ranger software with interactive visualization on the Loupe genome browser.

### HD haplotyping

The definitions of the most common 16 haplotypes in HD subjects were described previously [25]. The here described HTT haplotypes were constructed using 20 SNPs and 1 insertion-deletion polymorphism (rs149109767) and were named based on the frequency in the HD subjects with European ancestry. The 16 most frequent haplotypes account for approximately 90% and 82% of disease and normal chromosomes in HD subjects with European ancestry, respectively. We defined HTT haplotypes further based on the 1000 Genomes Project (KGP) data (<https://www.internationalgenome.org/phase3> data). Briefly, we extracted the same haplotype-defining 21 variations from all of the phased KGP data (5008 chromosomes) and constructed haplotypes subsequently. We then calculated the frequency in KGP data (all populations) for unique haplotype. Excluding the 16 haplotypes, which we had defined previously [25], we sorted haplotypes based on frequency and assigned names accordingly. Finally, 253 additional haplotypes were defined. The full list of the newly defined HD haplotypes is available in the supplementary information. HTT haplotypes were clustered via binary distance. A dendrogram was generated using Ward's linkage. This analysis was performed separately for the normal and mutant haplotypes. In each dendrogram, three clusters were identified via k-means clustering represented by different colors. Analysis was performed in R using packages *circlize* [74] and *dendextend* [75]. Barplots representing haplotype frequencies were generated via *ggplot2* [76].

### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13059-022-02752-5>.

Additional file 1. Metadata; All identifiers and demographic and clinical data.

Additional file 2. RNAseq adipose data file; Outliers removed and normalized counts.

Additional file 3. RNAseq muscle data file; Outliers removed and normalized counts.

Additional file 4. RNAseq adipose analysis file; DESeq2 analysis results.

Additional file 5. RNAseq muscle analysis file; DESeq2 analysis results.

Additional file 6. RNAseq fibroblasts data file; Outliers removed and normalized counts.

Additional file 7. RNAseq fibroblasts analysis file; DESeq2 analysis results.

Additional file 8. Proteomics adipose peptide datafile; Normalized LFQ values.

Additional file 9. Proteomics adipose protein datafile; Normalized LFQ values.

Additional file 10. Proteomics muscle peptide datafile; Normalized LFQ values.

Additional file 11. Proteomics muscle protein datafile; Normalized LFQ values.

Additional file 12. Proteomics skin peptide datafile; Normalized LFQ values.

Additional file 13. Proteomics skin protein datafile; Normalized LFQ values.

Additional file 14. Proteomics adipose analysis file; ROTS analysis results.

Additional file 15. Proteomics adipose analysis file; ROTS analysis results.

Additional file 16. Proteomics muscle analysis file; ROTS analysis results.

Additional file 17. Proteomics muscle analysis file; ROTS analysis results.

Additional file 18. Proteomics skin peptide analysis file; ROTS analysis results.

Additional file 19. Proteomics skin protein analysis file; ROTS analysis results.

Additional file 20. HTT SNP haplotyping fibroblasts; HTT centric SNP genotyping data.

Additional file 21. HD haplotype definitions; Novel HD haplotype definitions.

Additional file 22. Additional Figure 1.

Additional file 23. Review History.

#### Acknowledgements

First and foremost, we thank all participants of the study for their time and their willingness to undergo tissue biopsies. The authors acknowledge support by the High Performance and Cloud Computing Group at the Zentrum für Datenverarbeitung of the University of Tübingen, the state of Baden-Württemberg through bwHPC, and the German Research Foundation (DFG) through grant no INST 37/935-1 FUGG.

#### Review history

The review history is available as Additional file 23.

#### Peer review history

Stephanie McClelland was the primary editor of this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

#### Authors' contributions

Conceptualization: S.J.T. and M.O.; methodology: A.N., M.O.; software: A.N.; validation: A.N., K.K., T.H., J.P., S.J.T., M.O.; formal analysis: A.N.; investigation: A.N., K.K., T.H., N.B., J.H., S.T., J.P., H.S., S.H., S.J.T., M.O.; resources: A.N., D.J.L., J.M.L., S.J.T., M.O.; data curation: A.N., M.F., M.O.; writing—original draft: A.N., M.O.; writing—review and editing: all authors; visualization: A.N.; supervision: S.J.T., M.O.; project administration: S.J.T., M.O. The authors read and approved the final manuscript.

#### Funding

CHDI foundation grant A-5497 (S.J.T., M.O.). Ministerium für Wissenschaft, Forschung und Kunst Baden-Württemberg grant BioDATEN, BW Science Data Center (A.N.). Deutsche Huntington-Hilfe e.V. (A.N.). German Research Foundation, DFG grant NE 2372/1-1 (A.N.). UK Dementia Research Institute, DRI Ltd. (S.J.T.). UK Medical Research Council (S.J.T.). Alzheimer's Society and Alzheimer's Research UK (S.J.T.). Wellcome Trust grant 200181/Z/15/Z (S.J.T.). National Institutes of Health grants NS091161, NS105709, NS119471 (J.M.L.).

#### Availability of data and materials

All analyses were conducted in R v3.5.2, v3.6.1 and v4.0.3 [77]. Further details about the bioinformatics evaluation, as well as scripts and code are available at the github repository [https://github.com/ANeueder/MTM-HD\\_data\\_paper](https://github.com/ANeueder/MTM-HD_data_paper) under a GPL3 license [78]. The datasets supporting the conclusions of this article are included within the article and its additional files. RNAseq raw data files (fastQ) can be accessed through the European Genome-phenome Archive (EGA) by applying for data use access through EGA. Accession numbers are as follows: Skeletal muscle RNAseq: EGAS00001006474, adipose tissue RNAseq: EGAS00001006473 and fibroblast RNAseq: EGAS00001006472. Proteomics raw data files are available from the CHDI foundation upon reasonable request ([darren.freeman@chdifoundation.org](mailto:darren.freeman@chdifoundation.org)).

#### Declarations

##### Ethics approval and consent to participate

The local ethics committees at Ulm University and University College London approved the study (Ulm: 265-12; London: 12/LO/1565), and written informed consent was obtained from each participant. All experimental methods comply with the Helsinki Declaration.

##### Consent for publication

Not applicable.

**Competing interests**

J.M.L. serves in the scientific advisory board of GenEdit, Inc. A.N. acts as a consultant for Triplet Therapeutics, Inc. In the past 2 years, through the offices of UCL Consultants Ltd, a wholly owned subsidiary of University College London, S.J.T. has undertaken consultancy services for Alnylam Pharmaceuticals Inc., Atalanta Pharmaceuticals, F. Hoffmann-La Roche Ltd, Genentech, Guidepoint, Horama, Locanobio, LoQus23 Therapeutics Ltd, Novartis Pharma, PTC Therapeutics, Sanofi, Spark Therapeutics, Takeda Pharmaceuticals Ltd, Triplet Therapeutics, University College Irvine and Vertex Pharmaceuticals Incorporated. All other authors declare no competing interests.

Received: 14 February 2022 Accepted: 18 August 2022

Published online: 07 September 2022

**References**

- Bates GP, Dorsey R, Gusella JF, Hayden MR, Kay C, Leavitt BR, et al. Huntington disease. *Nat Rev Dis Primers*. 2015;1:15005.
- Carroll JB, Bates GP, Steffan J, Saft C, Tabrizi SJ. Treating the whole body in Huntington's disease. *Lancet Neurol*. 2015;14:1135–42.
- Cattaneo E, Zuccato C, Tartari M. Normal huntingtin function: an alternative approach to Huntington's disease. *Nat Rev Neurosci*. 2005;6:919–30.
- Saudou F, Humbert S. The Biology of Huntingtin. *Neuron*. 2016;89:910–26.
- Lu AT, Narayan P, Grant MJ, Langfelder P, Wang N, Kwak S, et al. DNA methylation study of Huntington's disease and motor progression in patients and in animal models. *Nat Commun*. 2020;11:4529.
- Morozko EL, Smith-Geater C, Monteys AM, Pradhan S, Lim RG, Langfelder P, et al. PIAS1 modulates striatal transcription, DNA damage repair, and SUMOylation with relevance to Huntington's disease. *Proc Natl Acad Sci U S A*. 2021;118:e2021836118.
- Steffan JS, Agrawal N, Pallos J, Rockabrand E, Trotman LC, Slepko N, et al. SUMO modification of Huntingtin and Huntington's disease pathology. *Science*. 2004;304:100–4.
- Ochaba J, Monteys AM, O'Rourke JG, Reidling JC, Steffan JS, Davidson BL, et al. PIAS1 regulates mutant huntingtin accumulation and Huntington's disease-associated phenotypes in vivo. *Neuron*. 2016;90:507–20.
- O'Rourke JG, Gareau JR, Ochaba J, Song W, Rasko T, Reverter D, et al. SUMO-2 and PIAS1 modulate insoluble mutant huntingtin protein accumulation. *Cell Rep*. 2013;4:362–75.
- Sedighi F, Adegbuyiro A, Legleiter J. SUMOylation prevents huntingtin fibrillization and localization onto lipid membranes. *ACS Chem Neurosci*. 2020;11:328–43.
- Ramirez-Jarquin UN, Sharma M, Zhou W, Shahani N, Subramaniam S. Deletion of SUMO1 attenuates behavioral and anatomical deficits by regulating autophagic activities in Huntington disease. *Proc Natl Acad Sci U S A*. 2022;119:e2107187119.
- de Souza JM, Abd-Elrahman KS, Ribeiro FM, Ferguson SSG. mGluR5 regulates REST/NRSF signaling through N-cadherin/beta-catenin complex in Huntington's disease. *Mol Brain*. 2020;13:118.
- Johnson R, Buckley NJ. Gene dysregulation in Huntington's disease: REST, microRNAs and beyond. *NeuroMolecular Med*. 2009;11:183–99.
- Ravache M, Weber C, Merienne K, Trottier Y. Transcriptional activation of REST by Sp1 in Huntington's disease models. *PLoS One*. 2010;5:e14311.
- Bondulich MK, Jolinon N, Osborne GF, Smith EJ, Rattray I, Neueder A, et al. Myostatin inhibition prevents skeletal muscle pathophysiology in Huntington's disease mice. *Sci Rep*. 2017;7:14275.
- von Maltzahn J, Jones AE, Parks RJ, Rudnicki MA. Pax7 is critical for the normal function of satellite cells in adult skeletal muscle. *Proc Natl Acad Sci U S A*. 2013;110:16474–9.
- Suomi T, Seyednasrollah F, Jaakkola MK, Faux T, Elo LL. ROTS: an R package for reproducibility-optimized statistical testing. *PLoS Comput Biol*. 2017;13:e1005562.
- Yuan Q, Li XD, Zhang SM, Wang HW, Wang YL. Extracellular vesicles in neurodegenerative diseases: insights and new perspectives. *Genes Dis*. 2021;8:124–32.
- La Spada AR, Wilson EM, Lubahn DB, Harding AE, Fischbeck KH. Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. *Nature*. 1991;352:77–9.
- David G, Abbas N, Stevanin G, Durr A, Yvert G, Cancel G, et al. Cloning of the SCA7 gene reveals a highly unstable CAG repeat expansion. *Nat Genet*. 1997;17:65–70.
- Wang G, Liu X, Gaertig MA, Li S, Li XJ. Ablation of huntingtin in adult neurons is nondeleterious but its depletion in young mice causes acute pancreatitis. *Proc Natl Acad Sci U S A*. 2016;113:3359–64.
- Kordasiewicz HB, Stanek LM, Wancewicz EV, Mazur C, McAlonis MM, Pytel KA, et al. Sustained therapeutic reversal of Huntington's disease by transient repression of huntingtin synthesis. *Neuron*. 2012;74:1031–44.
- Barker RA, Fujimaki M, Rogers P, Rubinsztein DC. Huntingtin-lowering strategies for Huntington's disease. *Expert Opin Investig Drugs*. 2020;29:1125–32.
- Marxreiter F, Stemick J, Kohl Z. Huntingtin lowering strategies. *Int J Mol Sci*. 2020;21:2146.
- Chao MJ, Gillis T, Atwal RS, Mysore JS, Arjomand J, Harold D, et al. Haplotype-based stratification of Huntington's disease. *Eur J Hum Genet*. 2017;25:1202–9.
- Kim KH, Abu Elneel K, Shin JW, Keum JW, Seong D, Kwak S, et al. Full sequence of mutant huntingtin 3'-untranslated region and modulation of its gene regulatory activity by endogenous microRNA. *J Hum Genet*. 2019;64:995–1004.
- Lee JM, Kim KH, Shin A, Chao MJ, Abu Elneel K, Gillis T, et al. Sequence-level analysis of the major European Huntington disease haplotype. *Am J Hum Genet*. 2015;97:435–44.
- Horvath S, Langfelder P, Kwak S, Aaronson J, Rosinski J, Vogt TF, et al. Huntington's disease accelerates epigenetic aging of human brain and disrupts DNA methylation levels. *Aging (Albany NY)*. 2016;8:1485–512.

29. Fusaki N, Ban H, Nishiyama A, Saeki K, Hasegawa M. Efficient induction of transgene-free human pluripotent stem cells using a vector based on Sendai virus, an RNA virus that does not integrate into the host genome. *Proc Jpn Acad Ser B Phys Biol Sci.* 2009;85:348–62.
30. Yan S, Li S, Li XJ. Use of large animal models to investigate Huntington's diseases. *Cell Regen.* 2019;8:9–11.
31. Rangel-Barajas C, Rebec GV. Overview of Huntington's disease models: neuropathological, molecular, and behavioral differences. *Curr Protoc Neurosci.* 2018;83:e47.
32. Neueder A, Orth M. Mitochondrial biology and the identification of biomarkers of Huntington's disease. *Neurodegener Dis Manag.* 2020;10:243–55.
33. Hodges A, Strand AD, Aragaki AK, Kuhn A, Sengstag T, Hughes G, et al. Regional and cellular gene expression changes in human Huntington's disease brain. *Hum Mol Genet.* 2006;15:965–77.
34. Labadorf AT, Myers RH. Evidence of extensive alternative splicing in post mortem human brain HTT transcription by mRNA sequencing. *PLoS One.* 2015;10:e0141298.
35. Wojtowicz S, Strosznajder AK, Jezyna M, Strosznajder JB. The novel role of PPAR alpha in the brain: promising target in therapy of Alzheimer's disease and other neurodegenerative disorders. *Neurochem Res.* 2020;45:972–88.
36. Altinoz MA, Ozpinar A, Ozpinar A, Hacker E. Erucic acid, a nutritional PPARdelta-ligand may influence Huntington's disease pathogenesis. *Metab Brain Dis.* 2020;35:1–9.
37. Dickey AS, Pineda VV, Tsunemi T, Liu PP, Miranda HC, Gilmore-Hall SK, et al. PPAR-delta is repressed in Huntington's disease, is required for normal neuronal function and can be targeted therapeutically. *Nat Med.* 2016;22:37–45.
38. Corona JC, Duchon MR. PPARgamma as a therapeutic target to rescue mitochondrial function in neurological disease. *Free Radic Biol Med.* 2016;100:153–63.
39. Chiang MC, Chern Y, Huang RN. PPARgamma rescue of the mitochondrial dysfunction in Huntington's disease. *Neurobiol Dis.* 2012;45:322–8.
40. Jin J, Albertz J, Guo Z, Peng Q, Rudow G, Troncoso JC, et al. Neuroprotective effects of PPAR-gamma agonist rosiglitazone in N171-82Q mouse model of Huntington's disease. *J Neurochem.* 2013;125:410–9.
41. UniProt C. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.* 2021;49:D480–9.
42. Fukuda M. TBC proteins: GAPs for mammalian small GTPase Rab? *Biosci Rep.* 2011;31:159–68.
43. Yuan W, Song C. The emerging role of Rab5 in membrane receptor trafficking and signaling pathways. *Biochem Res Int.* 2020;2020:4186308.
44. Pal A, Severin F, Lommer B, Shevchenko A, Zerial M. Huntingtin-HAP40 complex is a novel Rab5 effector that regulates early endosome motility and is up-regulated in Huntington's disease. *J Cell Biol.* 2006;172:605–18.
45. Kegel KB, Kim M, Sapp E, McIntyre C, Castano JG, Aronin N, et al. Huntingtin expression stimulates endosomal-lysosomal activity, endosome tubulation, and autophagy. *J Neurosci.* 2000;20:7268–78.
46. Trajkovic K, Jeong H, Krainc D. Mutant huntingtin is secreted via a late endosomal/lysosomal unconventional secretory pathway. *J Neurosci.* 2017;37:9000–12.
47. Mathews PM, Levy E. Exosome production is key to neuronal endosomal pathway integrity in neurodegenerative diseases. *Front Neurosci.* 2019;13:1347.
48. Lizarraga-Valderrama LR, Sheridan GK. Extracellular vesicles and intercellular communication in the central nervous system. *FEBS Lett.* 2021;595:1391–410.
49. Yates AG, Pink RC, Erdbrugger U, Siljander PR, Dellar ER, Pantazi P, et al. In sickness and in health: the functional role of extracellular vesicles in physiology and pathology in vivo: Part II: Pathology: Part II: Pathology. *J Extracell Vesicles.* 2022;11:e12190.
50. Sanchez II, Nguyen TB, England WE, Lim RG, Vu AQ, Miramontes R, et al. Huntington's disease mice and human brain tissue exhibit increased G3BP1 granules and TDP43 mislocalization. *J Clin Invest.* 2021;131:e140723.
51. Landwehrmeyer GB, Fitzer-Attas CJ, Giuliano JD, Goncalves N, Anderson KE, Cardoso F, et al. Data analytics from enroll-HD, a global clinical research platform for Huntington's disease. *Mov Disord Clin Pract.* 2017;4:212–24.
52. Tabrizi SJ, Ghosh R, Leavitt BR. Huntingtin lowering strategies for disease modification in Huntington's disease. *Neuron.* 2019;101:801.
53. Chang CY, Ting HC, Liu CA, Su HL, Chiou TW, Lin SZ, et al. Induced pluripotent stem cell (iPSC)-based neurodegenerative disease models for phenotype recapitulation and drug screening. *Molecules.* 2020;25:2000.
54. Csobonyeiova M, Polak S, Danisovic L. Recent overview of the use of iPSCs Huntington's disease modeling and therapy. *Int J Mol Sci.* 2020;21:2239.
55. Orth M, Gregory S, Scahill RI, Mayer IS, Minkova L, Kloppel S, et al. Natural variation in sensory-motor white matter organization influences manifestations of Huntington's disease. *Hum Brain Mapp.* 2016;37:4615–28.
56. Tabrizi SJ, Langbehn DR, Leavitt BR, Roos RA, Durr A, Craufurd D, et al. Biological and clinical manifestations of Huntington's disease in the longitudinal TRACK-HD study: cross-sectional analysis of baseline data. *Lancet Neurol.* 2009;8:791–801.
57. Huntington Study Group. Unified Huntington's Disease Rating Scale: reliability and consistency. *Huntington Study Group. Mov Disord.* 1996;11:136–42.
58. Penney JB Jr, Vonsattel JP, MacDonald ME, Gusella JF, Myers RH. CAG repeat number governs the development rate of pathology in Huntington's disease. *Ann Neurol.* 1997;41:689–92.
59. Orth M. European Huntington's Disease N, Handley OJ, Schwenke C, Dunnett S, Wild EJ, Tabrizi SJ, Landwehrmeyer GB: Observing Huntington's disease: the European Huntington's Disease Network's REGISTRY. *J Neurol Neurosurg Psychiatry.* 2011;82:1409–12.
60. Weiss A, Trager U, Wild EJ, Grueninger S, Farmer R, Landles C, et al. Mutant huntingtin fragmentation in immune cells tracks Huntington's disease progression. *J Clin Invest.* 2012;122:3731–6.
61. Demestre M, Orth M, Fohr KJ, Achberger K, Ludolph AC, Liebau S, et al. Formation and characterisation of neuromuscular junctions between hiPSC derived motoneurons and myotubes. *Stem Cell Res.* 2015;15:328–36.
62. Andrews S: FastQC: a quality control tool for high throughput sequence data. 2010. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
63. 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.

64. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods*. 2017;14:417–9.
65. Oldham MC, Langfelder P, Horvath S. Network methods for describing sample relationships in genomic datasets: application to Huntington's disease. *BMC Syst Biol*. 2012;6:63.
66. Soneson C, Love MI, Robinson MD. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Res*. 2015;4:1521.
67. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15:550.
68. Stephens M. False discovery rates: a new deal. *Biostatistics*. 2017;18:275–94.
69. Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics*. 2013;14:128.
70. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res*. 2016;44:W90–7.
71. Zhang Z. Missing data exploration: highlighting graphical presentation of missing pattern. *Ann Transl Med*. 2015;3:356.
72. 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.
73. Miller JA, Cai C, Langfelder P, Geschwind DH, Kurian SM, Salomon DR, et al. Strategies for aggregating gene expression data: the collapseRows R function. *BMC Bioinformatics*. 2011;12:322.
74. Gu Z, Gu L, Eils R, Schlesner M, Brors B. circlize implements and enhances circular visualization in R. *Bioinformatics*. 2014;30:2811–2.
75. Galili T. dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. *Bioinformatics*. 2015;31:3718–20.
76. Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, et al. Welcome to the Tidyverse. *J Open Source Software*. 2019;4:1686.
77. Team RC: R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2022.
78. Neueder A: Bioinformatics scripts for the MTM-HD paper. Github. 2022. [https://github.com/ANeueder/MTM-HD\\_data\\_paper](https://github.com/ANeueder/MTM-HD_data_paper).

### Publisher's Note

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

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

