Huntington disease gene expression signatures in blood compared to brain of YAC128 mice as candidates for monitoring of pathology

Kuijper, E.C.; Toonen, L.J.A.; Overzier, M.; Tsonaka, R.; Hettne, K.; Roos, M.; ... ; Mina, E.

Molecular Neurobiology (2022) 59:2532–2551
https://doi.org/10.1007/s12035-021-02680-8

Huntington Disease Gene Expression Signatures in Blood Compared to Brain of YAC128 Mice as Candidates for Monitoring of Pathology

Elsa C. Kuijper[1] · Lodewijk J. A. Toonen[1] · Maurice Overzier[1] · Roula Tsonaka[2] · Kristina Hettne[1] · Marco Roos[1] · Willeke M. C. van Roon-Mom[1] · Eleni Mina[1]

Received: 17 May 2021 / Accepted: 3 December 2021 / Published online: 29 January 2022

Abstract

While the genetic cause of Huntington disease (HD) is known since 1993, still no cure exists. Therapeutic development would benefit from a method to monitor disease progression and treatment efficacy, ideally using blood biomarkers. Previously, HD-specific signatures were identified in human blood representing signatures in human brain, showing biomarker potential. Since drug candidates are generally first screened in rodent models, we aimed to identify HD signatures in blood and brain of YAC128 HD mice and compare these with previously identified human signatures. RNA sequencing was performed on blood withdrawn at two time points and four brain regions from YAC128 and control mice. Weighted gene co-expression network analysis was used to identify clusters of co-expressed genes (modules) associated with the HD genotype. These HD-associated modules were annotated via text-mining to determine the biological processes they represented. Subsequently, the processes from mouse blood were compared with mouse brain, showing substantial overlap, including protein modification, cell cycle, RNA splicing, nuclear transport, and vesicle-mediated transport. Moreover, the disease-associated processes shared between mouse blood and brain were highly comparable to those previously identified in human blood and brain. In addition, we identified HD blood-specific pathology, confirming previous findings for peripheral pathology in blood. Finally, we identified hub genes for HD-associated blood modules and proposed a strategy for gene selection for development of a disease progression monitoring panel.

Keywords Huntington disease · YAC128 mice · Network analysis · Gene expression · Biomarker

Abbreviations

CNS	central nervous system
CSF	cerebrospinal fluid
CPA	concept profile analysis
DE	differentially expressed
DGE	differential gene expression
FDR	false discovery rate
GO-BP	Gene Ontology biological process
HD	Huntington's disease
HTT	Huntingtin
logCPM	log counts per million
logFC	log fold change
mHTT	mutant HTT
NfL	neurofilament light
PCA	principal component analysis
T1	blood time point 1
T2	blood time point 2
WGCNA	weighted gene co-expression network analysis
WT	wild-type

Introduction

Huntington disease (HD) is a neurodegenerative disorder characterized by motor, cognitive, and psychiatric symptoms [1]. Neurodegeneration starts in the striatum, but eventually affects the whole brain. This autosomal dominant disease is caused by an expanded CAG repeat in the huntingtin gene (*HTT*). Each CAG triplet in the DNA codes for a glutamine amino acid in the corresponding protein. Whereas *HTT* normally contains 6 to 35 CAG repeats, a CAG repeat length of 40 or more causes HD [1]. Reduced penetrance of HD disease symptoms is seen for a CAG repeat range from 36

✉ Elsa C. Kuijper
e.c.kuijper@lumc.nl

[1] Department of Human Genetics, Leiden University Medical Center, 2333, ZC, Leiden, The Netherlands

[2] Department of Biomedical Data Sciences, Leiden University Medical Center, 2333, ZC, Leiden, The Netherlands

to 39. The subsequent expansion of a stretch of glutamine amino acids in the HTT protein causes loss of its normal function and toxic gain of function, leading to protein aggregation [2].

The HTT protein interacts with numerous other proteins and has many physiological functions. For example, HTT is thought to be involved in transcriptional regulation, such as promoting brain-derived neurotrophic factor (BDNF) expression, synaptic activity, anti-apoptotic activity, and nucleocytoplasmic shuttling [1]. The HTT protein is essential during development and remains widely expressed thereafter. Highest HTT protein levels are found in the central nervous system (CNS), followed by the stomach, testes, and bone marrow [3].

Many pathophysiological mechanisms caused by mutant HTT (mHTT) have been described since the discovery of the genetic cause of HD in 1993, but there is still no cure [4, 5]. Only symptomatic treatment is available so far, mainly targeting neurochemical systems and addressing hyperkinesia. However, promising trials are ongoing at the moment that target the cause of HD and aim to lower mHTT expression [5, 6]. The development of any HD therapy would benefit from methods that can monitor disease progression and response to treatment.

Given the complex pathogenesis of HD, which involves various pathogenic pathways and non-specific secondary disease processes, a broad set of biomarkers is likely needed to provide complementary and comprehensive information about the state and progression of disease. Up to now, both imaging biomarkers and biomarkers in biofluids have been described [7]. Imaging techniques have shown to be informative preclinically, but these are relatively expensive and difficult to adequately automate [7, 8]. In cerebrospinal fluid (CSF), levels of mHTT can provide important pharmacodynamic information in huntingtin-lowering therapies. Moreover, neurofilament light (NfL) levels in CSF seem to reflect neuronal damage and show strong correlation with NfL levels measured in blood [7]. Blood biomarkers are desired as they are less invasive compared to CSF measurements and have relatively low costs. Nonetheless, as NfL reflects the final common pathophysiologic pathway of HD, biomarkers involved in more upstream processes might be useful as well.

Peripheral tissue such as blood could be suitable for disease monitoring as HTT expression is ubiquitous and HD involves peripheral symptoms. However, finding blood biomarkers is challenged by changes in blood reflecting non-specific systemic processes rather than neuropathology in HD [7, 9]. To face this challenge, previous research by Mina et al. focused on similarity between functional signatures linked to the HD phenotype present in both brain and blood of HD patients to enhance the chance of identifying disease-specific signatures in blood [8]. To identify disease signatures, Mina et al. used weighted gene co-expression network analysis (WGCNA). WGCNA detects gene sets with similar expression patterns (modules) in transcriptomics data. It is assumed that the genes within these modules function together in a biological process. Mina et al. identified and compared signatures at the level of biological processes across blood and brain. The benefit of examining functional signatures is emphasized by the fact that the expression of individual genes is poorly preserved between brain and blood [10], while similar pathways might be affected in blood and brain through different genes. Indeed, HD-specific functional signatures in human blood were shown to adequately represent the signatures in HD brain and hence could offer potential to be used as biomarkers [8].

As mouse models are commonly used in preclinical studies and studies into pathogenic mechanisms, translational biomarkers are desired to monitor disease and therapeutic efficacy in these models. Hence, in this study, we aimed to identify and compare HD disease signatures in HD mouse blood and brain. To that end, RNAseq data was collected from blood and brain from YAC128 HD mice followed until on average 20 months of age presenting an advanced stage of disease. The YAC128 model contains the full human *HTT* gene with 128 CAG repeats and demonstrates a motor phenotype and neuropathology [11]. An intermediate time point of blood sampling around 13 months of age, when YAC128 mice start to show striatal neuronal loss [11], was included to assess disease progression. We applied WGCNA and identified modules associated with HD that we annotated via text-mining to assess their biological function and compare functional signatures between blood and brain. We compared our findings with the disease signatures that were found in HD patients by Mina et al. [8] and prioritized genes for further investigation of their use in monitoring of disease.

Methods

Mice: Behavioral Testing and Tissue Collection

YAC128 transgenic HD mice maintained on the FVB/N strain [11] and wild-type FVB/N mice (WT) were obtained from Jackson Laboratories (Bar Harbor, Maine, USA) and bred in-house. All animal experiments were performed according to European Communities Council Directive 2010/63/EU and were approved by the Leiden University animal ethical committee (AVD1160020171069). Only male mice were used in the experiment to reduce variation. Wild-type and YAC128 mice were matched for age, which was on average 7 months of age at the start of the study (Supplementary Table 1). We assessed body weight and performed beamwalk and rotarod tests in order to assess the motor phenotype of the mice. For the beamwalk

test, mice had to cross a plastic cylindrical bar (10 or 30 mm diameter) with length of 80 cm connected between 2 boxes (20 × 20 × 20 cm) elevated at 53 cm height. Mice were placed in the transparent box to cross the bar to an enclosed dark box. The average latency to cross, number of drops, and hind slips for 3 trials per testing day were reported. The rotarod test included 3 trials per testing day, where mice were placed on a rotarod accelerating from 5 to 45 rpm. The maximum trial length was 500 s and the latency to fall was recorded. Statistical analysis was performed in Graph Pad Prism 8, assessing the differences between WT and YAC128 mice by the two-stage linear step-up multiple testing procedure of Benjamini, Krieger and Yekutieli with a Q of 5%.

Blood samples were taken for RNA sequencing in the morning after 6 months (time point 1, T1) from tail vein and before sacrifice after 14 months (time point 2, T2) from tail vein and orbital sinus (for time line of the study, see Fig. 1). Blood was collected in RNAprotect animal blood tubes (Qiagen) following the manufacturer's instructions, stored overnight at 4°C and subsequently frozen at −80°C until RNA isolation. Upon sacrifice, brainstem, cerebellum, striatum, and cortex were dissected, snap-frozen in liquid nitrogen, and stored at −80°C. At the start of the study, 8 mice per group were included. However, two transgenic mice deceased before the end of the study and blood sampling failed for one transgenic mouse at T2 (Table 1). The samples obtained from these mice were used for RNA sequencing and technical validation with qPCR and western blot.

RNA sequencing

RNA isolation, library preparation, sequencing, and sequence data processing were performed according to the procedure described by Toonen et al. [12]. In short, RNA was isolated from blood and brain using the RNeasy protect animal blood kit (Qiagen, Hilden, Germany) and PureLink RNA mini column kit (Thermo Fisher Scientific) respectively according to the manufacturer's instructions for total RNA isolation including DNAse treatment. Reduction of alpha and beta globin mRNA for RNA isolated from blood was performed using the GLOBINclear magnetic bead kit for mouse/rat (Qiagen) following the manufacturer's instructions. RNA was shipped to DeCode Genetics (Reykjavik, Iceland) where library preparation and RNA sequencing was performed. Quality of RNA was assessed with the LabChip GX using the 96-well RNA kit (Perkin Elmer). Non-strand-specific sample preparation was performed using the TruSeq Poly-A v2 kit (Illumina, San Diego, USA) following the manufacturer's instructions. Primary processing and base calling was performed with Illumina's HCS and RTA software. Demultiplexing and generation of FASTQ files was done with Illumina scripts (bcl2fastq v1.8). The FASTQ files for the mouse brain RNA can be found in the GEO repository, accession GSE107958, and blood samples are listed under accession GSE108069. Sequencing data was analyzed using the BIOPET Gentrap in-house pipeline (http://biopet-docs.readthedocs.io/en/v0.7.0/pipelines/gentrap/). The fastqc toolkit (v0.11.2) was used to evaluate sequencing quality (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Reads were cleaned with Sickle (v1.33 with default settings) and Cutadapt (v1.10, with default settings except for "-m 20") and aligned to the mouse reference genome build 10 (GRCm38/mm10) using STAR aligner version 2.3.0e (non-default settings used by STAR: "--outFilterMultimapNmax 1 –outFilterMismatchNmax 10 –outSJfilterReads Unique").

Differential Gene Expression Analysis

Differential gene expression (DGE) analysis was performed between WT and YAC128 for each time point in blood and each brain region separately using the edgeR package in R [13, 14]. Prior to DGE analysis, we filtered out genes with less than 4 counts per million (CPM) across all samples. In addition, we applied a GC-content correction using the conditional quantile normalization (CQN) package to remove bias related to the GC content in the samples of both brain and blood tissue [15]. We calculated and included the offset obtained from the GC-content correction when estimating dispersion in edgeR. Regarding the DGE in each tissue, we included as contrasts for the likelihood ratio test each brain region (brainstem, cerebellum, cortex, striatum) and each time point (T1, T2). We considered genes with a false discovery rate (FDR, Benjamini-Hochberg [16]) below 0.05 as significantly differentially expressed. Graphs were made in GraphPad Prism 8.

RT-qPCR validation

Selection of differentially expressed (DE) genes for technical validation of our RNAseq results was based on the FDR, fold change, and expression levels. cDNA was synthesized from RNA of blood, cerebellum, and striatum using the Transcriptor First Strand cDNA Synthesis Kit (Roche, Mannheim, Germany). We followed the manufacturer's protocol, using random hexamer primers and incubating the RT reaction for 10 min at 25°C followed by 30 min at 55°C. qPCR was performed with the SensiMix SYBR Hi-ROX Kit (Bioline, London, UK), using 3 µL 10× diluted cDNA, 250 nM forward and reverse primer, and 5 µL 2× SensiMix™ SYBR® in a total volume of 10 µL. The primers, designed with Primer3Plus [17], are listed in Table 2. The qPCR program, executed on the LightCycler480 (Roche), started with

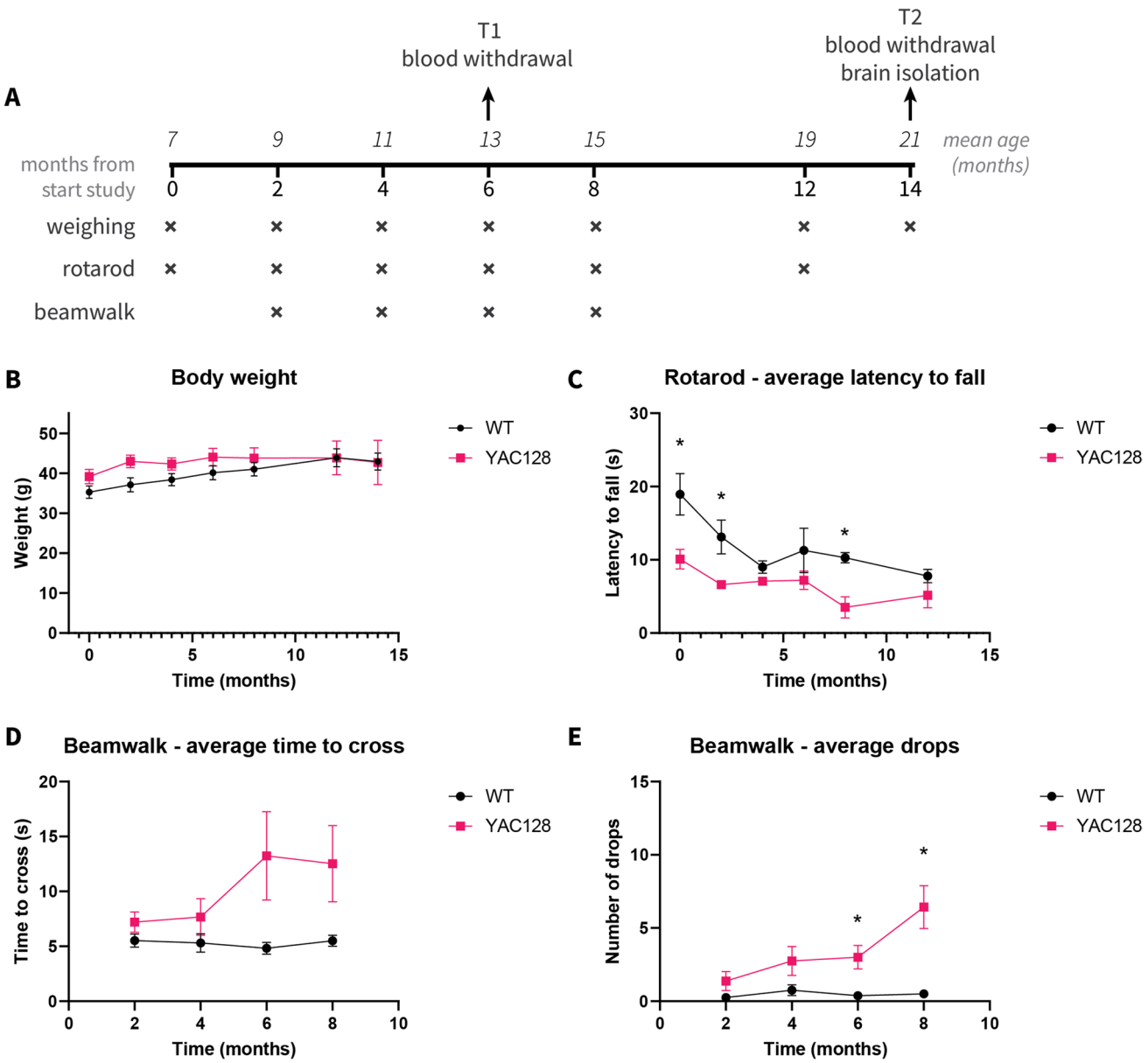


Fig. 1 Motor phenotype was confirmed in YAC128 mice. **A** During 14 months after start of the study, weighing and behavioral tests were performed. At 6 months (T1) and 14 months (T2), blood was withdrawn and after sacrifice brain material was isolated. **B** Weight was not significantly different between WT and YAC128 mice; **C** the average latency to fall from the rotarod was lower for YAC128 mice than WT mice, significant at 2 and 8 months; **D** the average time needed to cross the beamwalk was not significantly different for YAC128 and WT mice despite showing a trend towards increased time needed for YAC128 mice; **E** YAC128 mice did show a higher number of drops from the beamwalk, significant after 6 and 8 months. (mean and SEM shown, * indicates adjusted $P<0.05$, n=6–8 mice per group)

10-min pre-incubation at 95°C, followed by 45 amplification cycles (15 s at 95°C, 15 s at 60°C, 15 s at 72°C, acquire) and ended with a melting curve from 65 to 97°C (0.11°C/s). Expression values were corrected for primer efficiencies, as calculated by LinRegPCR [18], and related to the geometric mean of three housekeeping genes (*Actb* and *Hprt* for blood, cerebellum, and striatum in addition to *Rpl22* for blood and cerebellum and *Rpl27* for striatum). Statistical testing for differential gene expression between WT and YAC128 was performed in GraphPad 8, using the two-stage linear step-up multiple testing procedure of Benjamini, Krieger, and Yekutieli [19], with Q=5% and without assuming a consistent SD.

Table 1 Sample numbers available within the RNAseq dataset for blood time point 1 and 2 and brain.

Tissue	Number of mice *WT*	Number of mice *YAC128*
Blood T1	8	7
Blood T2	8	5
Brain (brainstem, cerebellum, cortex, striatum)	8	6

T1 time point 1, *T2* time point 2

Protein Isolation

For protein isolation, snap frozen cerebellar and striatal tissue was homogenized in RIPA buffer (50 mM Tris-HCl pH 8.0, 150 mM NaCl, 1% IGEPAL CA-630, 0.5% DOC, 0.1% SDS) including protease inhibitor (1 tablet per 10 ml, Roche) with use of the bullet blender Storm 24 (Next Advance, Troy, USA). Protein concentration was determined using the Pierce BCA Protein Assay Kit (ThermoScientific, Rockford, USA) according to the manufacturer's instructions. For storage at −20°C, glycerol was added to an end concentration of 5% and samples were snap frozen.

Western Blotting

Protein lysates were diluted to reach the indicated amount of protein and 4× Bolt LDS sample buffer (ThermoScientific) and 10× Bolt reducing agent were added. After incubating for 10 min at 70°C, samples were loaded onto a precast 10% Bis-Tris Plus SDS page gel (ThermoScientific) next to a prestained protein ladder (#26617, ThermoScientific) and run for approximately 1 h at 100–150V in MES running buffer (ThermoScientific). Subsequently, proteins were transferred to a nitrocellulose membrane with the Trans-Blot Turbo Transfer System (Bio-Rad, Veenendaal, the Netherlands) using the 1.5 mm gel protocol (1.3 A, 25 V, 10 min). Membranes were blocked in 5% milk/TBS for 1 h at RT. Primary antibodies (Table 3) were incubated o/n at 4°C or for 3 h at RT. After washing with TBS, secondary antibodies (Table 3) were incubated at RT for 1 h, followed by another wash with TBS. Detection was performed with the Odyssey infrared imaging system (LI-COR, Lincoln, USA). For quantification, signal intensity was determined in Image Studio Lite Version 5.2 (LI-COR) after background correction. To correct for differences in protein loading, signal intensities relative to β-actin signal were calculated. GraphPad 8 was used to determine statistically significant differences between the WT and YAC128 group, applying the two-stage linear step-up multiple testing procedure of Benjamini, Krieger, and Yekutieli [19], with Q=5% and without assuming a consistent SD.

Weighted Gene Co-expression Network Analysis

Prior to weighted gene co-expression network analysis (WGCNA), we performed principal component analysis (PCA) and hierarchical sample clustering to evaluate sample consistency and detect outliers within the different groups. In addition, we applied sample network analysis to remove sample outliers as described by Oldham et al. [20]. Two rounds of analysis were carried out, in which the samples with a connectivity Z-score lower than −2 were removed. An

Table 2 List of primers used for validation of RNAseq data with qPCR (5′-3′).

Gene	Forward primer	Reverse primer
Blood		
Oas2	GCATGAACATGCCCTTGTAG	AGCTGGGATTCTCATTGGAG
Oas3	CGCTAAACATCACCCTACAGC	AGTCGAGGAAGATGACGAGTTC
Cd226	CTTTACAGATGTCGCTCAGAGG	CAGTGAAACTAACCCTCCAACG
Rtp4	GGAGCCTGCATTTGGATAAG	TTCTGCAGCATCTGGAACAC
Cd4	TTCACCTGGAAGTTCTCTGACC	AACGATCCTTTCTCCCATGC
Brain		
Car2	ATAAAGCTGCGTCCAAGAGC	AGCCCCAGTGAAAGTGAAAC
Il33	GTCCCGCCTTGCAAAATAAG	TTATGGTGAGGCCAGAACG
Igfbp5	GGGTTTGCCTCAACGAAAAG	TAGGTCTCTTCAGCCATCTCG
Crhr1	AGCAGTGTGAGAGCCTGTCC	AGCGGACACCGTAGAAAAAG
Gabra5	ATGACCCAAACCCTCCTTG	CAAGAGTCCGTCCAAGATCC
Reference genes		
Actb	GGCTGTATTCCCCTCCATCG	CCAGTTGGTAACAATGCCATGT
Hprt	TCCCTGGTTAAGCAGTACAGCC	CGAGAGGTCCTTTTCACCAGC
Rpl22	AGGAGTCGTGACCATCGAAC	TTTGGAGAAAGGCACCTCTG
Rpl27	AAAGCCGTCATCGTGAAGAAC	GCTGTCACTTTCCGGGGATAG

exception was made for the cortex, where two samples were kept in our analysis despite a *Z*-score slightly below −2, because the hierarchical clustering dendrogram and PCA did not show obvious deviation from the other samples.

We applied the WGCNA algorithm (version 1.63) [21] in R to construct a weighted gene co-expression network for each data set separately. In order to meet the scale-free topology assumption ($R^2 > 0.8$) and obtain modules with comparable sizes for all regions, the soft thresholds and merging module thresholds for each network were respectively chosen as follows: blood T1—16, 0.0001; blood T2—16, 0.1; brainstem—16, 0.1; cerebellum—10, 0.25; cortex—14, 0.0001; striatum—18, 0.12. For all networks, the following parameters were chosen: minimum module size = 30 and method = "signed hybrid" regarding the type of network with deepSplit = 2.

To assess disease association of the modules from each network, we determined the correlation between each module eigengene and the genotype (WT or YAC128) and the significance of the correlation by Pearson correlation as implemented in the WGCNA package. Modules with correlation values >|0.5| and a *P*-value <0.05 were considered significant.

We performed WGCNA-based module preservation analysis to assess whether the HD-associated modules were preserved in the networks constituted for the individual brain regions as described before [22]. Module preservation statistics were calculated using networkType = "signed hybrid".

Module Annotation

In order to avoid potential bias from poorly annotated mouse genes during module annotation, we first mapped the mouse gene identifiers to their equivalent human orthologues using the Ensembl BioMart package (version 2.34.2; GRCm38 and GRCh38 genome assemblies) [20]. Details considering the frequencies of missing human orthologues and one-to-many human orthologues were documented (Supplementary Fig. 1).

We performed concept profile analysis (CPA) to annotate the modules identified by WGCNA by applying the workflows developed by Mina et al. [8]. CPA is a functional analysis technique based on literature mining [23, 24]. The top 20 of Gene Ontology biological process (GO-BP) annotations was selected for each module (Fig. 2). We grouped different GO-BP module annotations to a more general, semantically related term according to a manually composed index (Supplementary File I). This enabled comparisons between modules at a biologically similar ontological level as the originally retrieved GO-BP annotations were reflecting different levels of the ontological tree. Annotation of three randomly chosen modules with the curated GO Slim database using the GOTermMapper [25] showed results in line with annotation with our manually composed index (Supplementary file II). However, differences were also present as our index of grouped annotations was more detailed since we aimed to keep as much information as possible while creating a biologically meaningful overview of processes represented by the different modules.

Assessment of Module Annotation Overlap

We used the grouped module annotations of the HD-associated modules to assess the overlap between (a) mouse blood and brain and (b) human and mouse common blood-brain signatures. The overlap between the grouped module annotations for all HD-associated modules per tissue combined was depicted by an online tool for Venn diagrams [26]. Labeled heat maps showing the pairwise overlap of grouped module annotations for modules for each comparison were created in R using the WGCNA package (version 1.63) [21].

The original module annotations obtained by CPA were used to statistically evaluate the module overlaps between mouse blood and brain. Module annotations from both blood time points were compared with the module annotations from each brain region and the number of overlapping annotations were counted for each comparison. To statistically assess the module similarity based on the number of

Table 3 Antibodies used for validation of RNAseq data with western blot

Primary antibodies				
Protein	*Antibody number*	*Dilution*	*Species*	*Company*
β-actin	Ab6276	1:5000	Mouse	Abcam, Cambridge, UK
Gabra5	Ab175195	1:500	Rabbit	Abcam
Car2	NB600-919-100 ug	1:2000	Rabbit	Novus Biologicals, Littleton, USA
Plp1	PA3-150	1:1000	Rabbit	ThermoScientific
Secondary antibodies				
Name	*Antibody number*	*Dilution*	*Species*	*Company*
IRDye 800CW Anti-Rabbit	926-32211	1:10.000	Goat	LI-COR
IRDye 680RD Anti-Mouse	926-68070	1:10.000	Goat	LI-COR

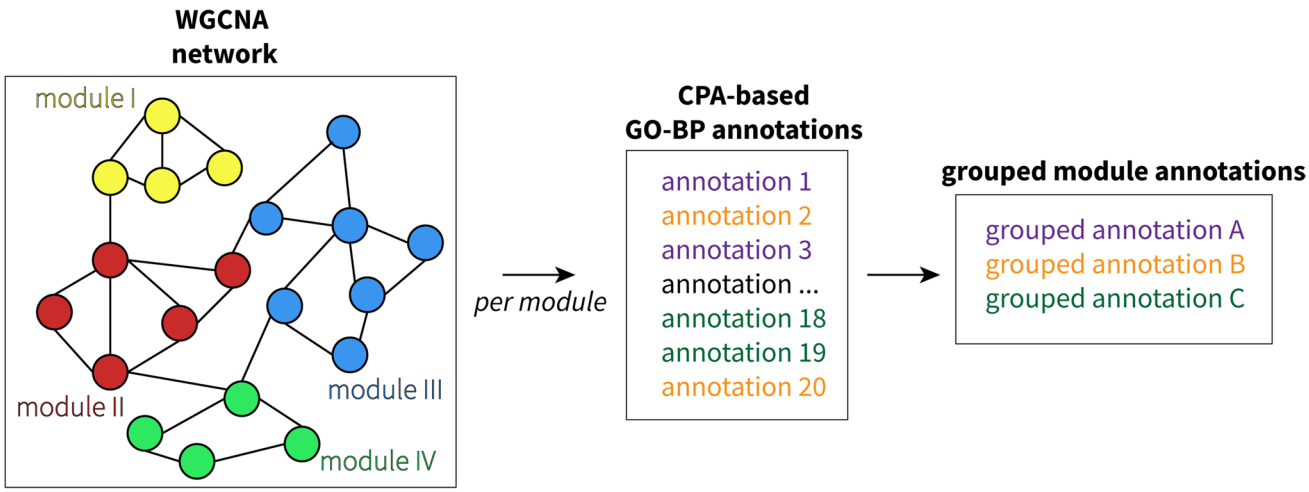


Fig. 2 Workflow for description of WGCNA modules using CPA-based annotations. WGCNA was used to identify modules of co-expressing genes. These modules were annotated with Gene Ontology terms of biological processes (GO-BP) using concept profile analysis (CPA). To compare the processes at a similar ontological level, the individual annotations were grouped based on a manually composed index

overlapping annotations, a randomization experiment was performed. For each module within a network, 100 random modules of the same number of genes were generated. Subsequently, these modules were also annotated as described above using CPA. Pairwise comparisons of the annotations from the random modules from different tissues were performed. To evaluate which original pairwise overlaps were significantly different from overlaps obtained from the random modules, the Westfall and Young's minP method [27] was used to assign *P*-values to the original module pairs. Module pairs were ranked based on the *P*-value and we assessed whether the blood modules predominantly overlapped with HD-associated or non-HD-associated brain modules.

Hub Gene Selection

For the blood modules showing an absolute correlation value with the HD genotype >0.65, hub genes were determined using the function for intramodular connectivity within the WGCNA package (version 1.63) in R. Hub genes were defined as the top 10% of intramodular connectivity (kWithin value) and subsequently filtered to meet the criteria of gene module membership >0.80 and gene significance >0.20. Based on the differential expression fold change and significance and level of expression of the hub genes, a selection of hub genes was made for which the expression as measured with RNAseq at T1 to T2 was assessed and linear regression analysis was performed in GraphPad 8 to compare the slopes between wild-type and YAC128 mice.

Results

Confirming the HD Phenotype in the YAC128 Mouse Model

Motor Phenotype Was Observed in YAC128 Mice

To confirm an HD phenotype in our YAC128 mice, we performed behavioral tests assessing motor deficits (Fig. 2A). During the study, body weight was not significantly different between YAC128 and control mice (Fig. 2B). YAC128 mice showed decreased latency to fall on the rotarod test compared to WT mice, being significant at the start of the study and after 2 and 8 months (Fig. 2C). The beamwalk test depicted a trend towards increased crossing time for YAC128 mice compared to WT mice, though this trend was not significant (Fig. 2D). The average number of drops increased over time for the YAC128 mice, becoming significantly different from WT mice after 6 months (Fig. 2E).

YAC128 Mice Exhibit an HD Gene Expression Phenotype in Both Blood and Brain

We applied DGE analysis to the RNAseq data from YAC128 and WT mice to evaluate pathology at gene expression level and observed an HD phenotype both in the blood and brain (Supplementary Table 2). Cerebellum and striatum exhibited the highest number of significantly differentially expressed (DE) genes (FDR < 0.05) among the different brain regions: 144 and 100 DE genes respectively. In brainstem, 59 genes

were differentially expressed and cortex exhibited 32 DE genes. In blood, there were 7 DE genes at time point T1, whereas 169 DE genes were detected at time point T2, reflecting progression of a blood disease phenotype in these YAC128 mice.

To technically validate the results of DGE analysis of the RNAseq data with independent techniques, we selected significant DE genes (FDR < 0.05) with a log fold change > |0.6| and high RNA expression levels for analysis with qPCR and western blot. For qPCR, we selected 5 genes differentially expressed in blood T2 and 5 genes that were differentially expressed in cerebellum and/or striatum. We confirmed a significant difference between YAC128 and WT expression levels for 4 of the 5 selected genes for blood T2 with qPCR (Fig. 3A, D), and for 1 out of 4 genes in cerebellum (Fig. 3B, E). For the genes that showed no significant difference, the direction of change was similar as observed with the DGE analysis of the RNAseq data and the spread of data points generally followed the same pattern. For striatum, both genes selected for validation showed significantly different expression levels with qPCR (Fig. 3C, F). Furthermore, we assessed 2 genes for striatum (*Car2* and *Plp1*) and 1 gene for cerebellum (*Gabra5*) on protein level using western blot (Fig. 3G–J, Supplementary Figs. 2A–B, 3). For *Gabra5*, we observed a difference on protein level in cerebellum consistent with the RNAseq data, but this was not significant as a large variation was observed which was in line with the RNAseq and qPCR data. The two genes selected for striatum showed significant differences in protein levels similar to the outcomes

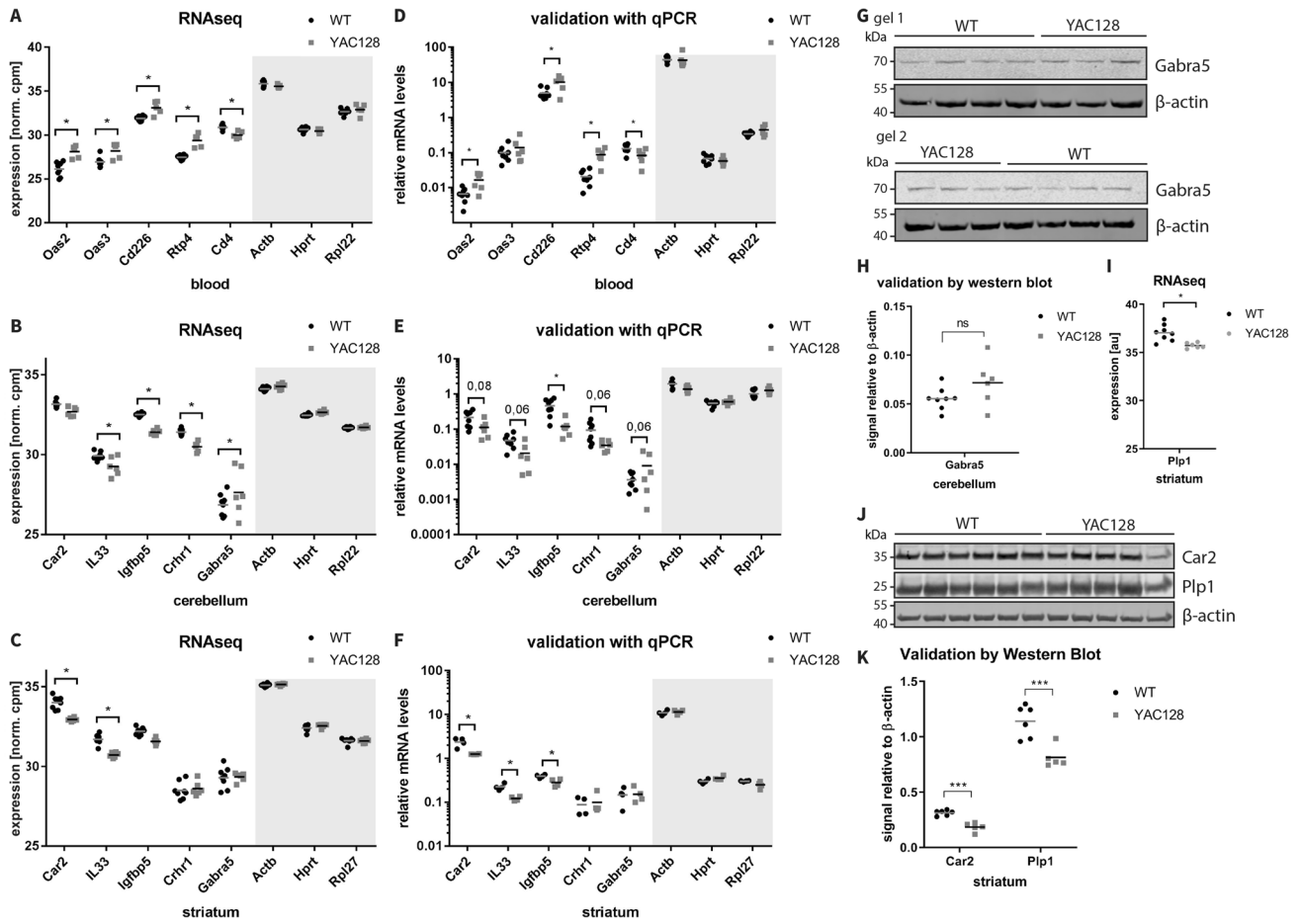


Fig. 3 Validation of RNAseq results with qPCR and western blot. **A–C** Expression levels measured by RNAseq in blood T2, cerebellum, and striatum of WT and YAC128 mice. **D–F** Expression levels measured with qPCR relative to the geometric mean of three reference genes (Actb, Hprt, Rpl22, or Rpl27) presented on log scale. Representative graphs of two independent experiments. Mean expression levels are listed in Supplementary Table 3. **G** Western blot signal for Gabra5 and housekeeping gene β-actin in cerebellum of 6 YAC128 mice and 8 WT mice (one blot); **I** RNAseq expression levels measured in striatum for additional genes selected for western blot validation; **J** Western blot signal for Car2, Plp1, and housekeeping gene β-actin in striatum of 5 YAC128 and 6 WT mice (representative images of two blots for Car2 and three blots for Plp1); **H**, **K** quantification of the Western blots shown in **G** and **J** respectively. (mean shown, * indicates adjusted $P<0.05$ and *** adjusted $P<0.001$, norm. cpm: normalized counts per million)

of DGE analysis. Even though qPCR and western blot are techniques that are less sensitive than RNAseq, challenging validation of genes with low expression levels such as *Gabra5*, we found corresponding changes showing the validity of our RNAseq data.

Functional Annotation of WGCNA Modules Revealed a Wide Range of Deregulated Processes in Both Brain and Blood of YAC128 Mice

To further describe the YAC128 phenotype, WGCNA was applied to each tissue separately to identify modules of co-expressed genes correlating with the HD genotype. Prior to performing WGCNA, the data were examined for sample outliers as these could easily affect outcomes when sample size is small. PCA showed a clear separation of the different brain regions and the two blood time points (Supplementary Fig. 4). However, two brainstem and two cerebellum samples were identified as outliers. This was confirmed by sample network analysis which identifies outliers based on network connectivity normalized to a *Z*-score (Supplementary Fig. 5). Similarly, for both blood T1 and T2, two samples were classified as outliers (Supplementary Fig. 6). Hence, these samples were removed and with the remaining number of samples (12 for brainstem, cerebellum, blood T1, and blood T2 and 14 for cortex and striatum), networks were created for each tissue consisting of clusters of co-expressed genes (modules) applying WGCNA. On average, the module size in the different networks was comparable with mean module size ranging from 153 to 233 genes per module (Supplementary Table 4).

Next, we assessed the modules that showed significant correlation ($P < 0.05$) with the HD genotype (Supplementary Figs. 7, 8). Assessing brain first, the highest number of HD-associated modules occurred in brainstem (17 modules), whereas only 7 modules from striatum were associated with the genotype. Cerebellum exhibited 11 HD-associated modules and cortex 10. Both positive and negative correlations of the module eigengenes with the HD genotype were found, indicating respectively overall upregulation or overall downregulation of gene expression within a module in YAC128 mice compared to WT mice.

To determine which biological processes are represented by the HD-associated modules, we examined the module annotations obtained by CPA (Fig. 4, Supplementary File III). Since these original annotations reflected different levels of the GO tree, we grouped several annotations under their most representative parental annotations ("grouped module annotations") to enable a functional comparison between the different regions (Figure 1, Supplementary File I). The grouped annotations of modules that showed the strongest upregulation ($r \geq 0.8$) in YAC128 brain mainly represented "cell cycle," "vesicle-mediated transport," "nuclear transport," "protein modification," and "proteolysis." "Glial cell differentiation" and "myelination" were grouped module annotations that occurred most frequently in the strongest downregulated ($r \geq 0.8$) modules of brainstem and striatum. Other grouped module annotations that occurred in HD-associated modules in brain were "RNA processing," "autophagy," "translation," "cellular component organization," "chromosome organization," and "histone modification." Although there were differences in the number of HD-associated modules, the grouped module annotations did not show major differences among all four brain regions indicating that eventually mHTT exerts similar effects across the brain at late-stage pathology in 20-month-old YAC128 mice.

In blood T1, 5 HD-associated modules were detected with WGCNA. The grouped module annotations of upregulated HD-associated modules included "leukocyte-mediated immunity," "cytokine activity," "viral immunity," and "protein modification" (Fig. 5A). Only one module was downregulated which represented the grouped module annotations "leukocyte-mediated immunity" and "nuclear transport." Interestingly, when looking at the ungrouped module annotations related to immunity of these blood T1 HD-associated modules, the upregulated modules mainly contained processes related to innate immunity such as macrophage and neutrophil activation whereas the annotations from the downregulated module included T cell immunity and antigen presentation (Supplementary File III–Blood T1), indicating that an innate immune response is evoked.

A higher number of HD-associated modules was observed in blood T2 (11 modules) compared to blood T1, pointing towards progression of disease that was observed in the DGE analysis as well. CPA-based grouped module annotations showed that the modules mainly represented "protein modification," "histone modification," "nuclear transport," "RNA processing," and "DNA repair" (Fig. 5B, Supplementary File III). In contrast to blood T1, the majority of HD-associated modules was downregulated in blood T2 (Fig. 5). Whereas blood T1 modules were mainly involved in immunity, only two modules represented leukocyte-mediated immunity in blood T2. The module *bisque4* showed annotations related to "leukocyte chemotaxis and activation." Interestingly, 10% of the genes (21 genes) within module *bisque4* occurred in blood T1 in module *darkred*, which was involved in immunity as well but upregulated instead of downregulated. This switch to downregulation is in line with the observation that immunity-related processes fades into the background at time point 2. Blood T2 module *pink* contained annotations mainly related to T cells and shared 18% of its genes (37 genes) with the T1 module *lightyellow*, which was downregulated as well. Furthermore, when we looked at the grouped

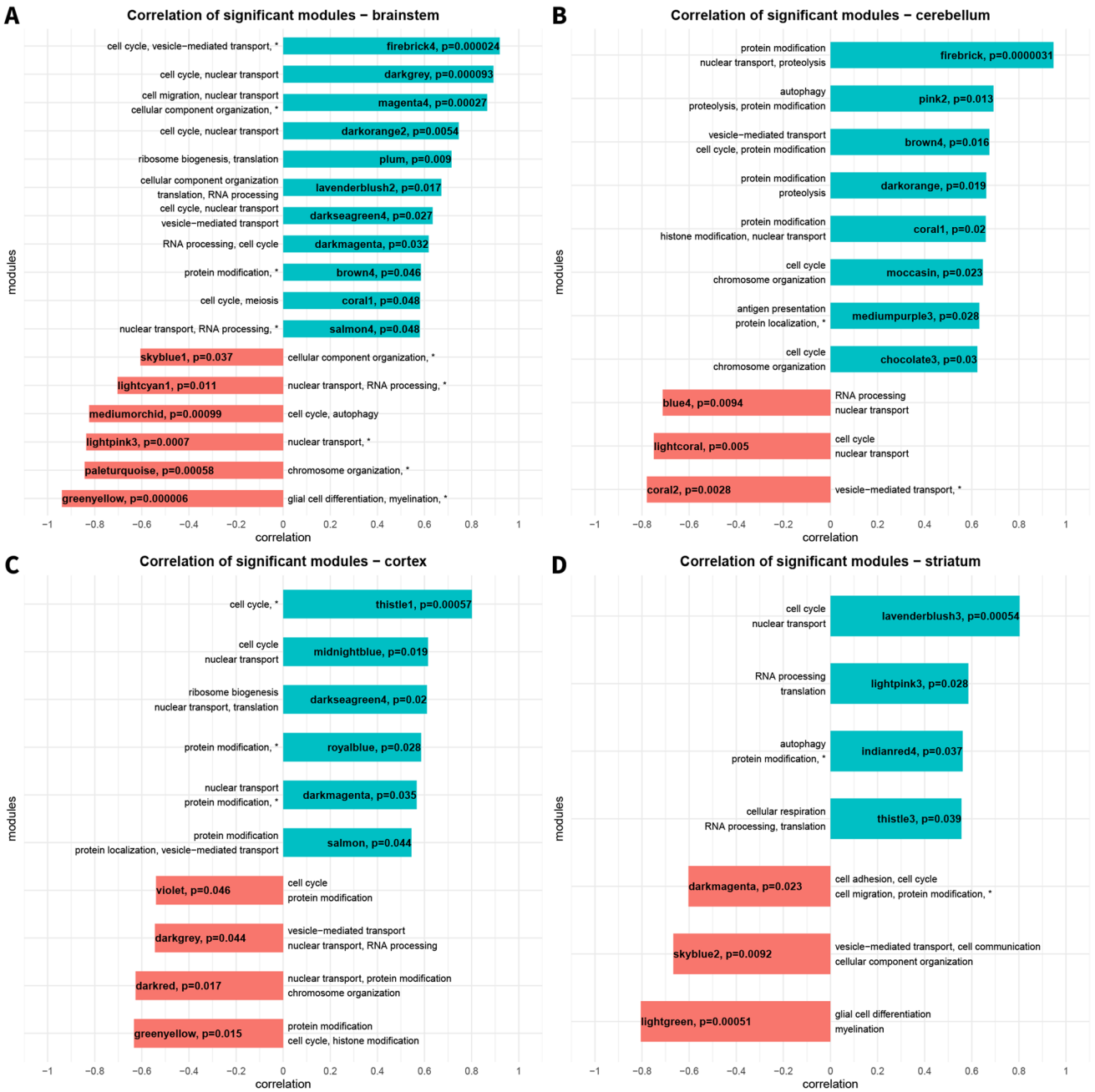


Fig. 4 Overview of significant HD-associated WGCNA modules in the different brain regions. For significant modules identified in brainstem (**A**), cerebellum (**B**), cortex (**C**), and striatum (**D**), the correlation with the HD genotype and its *P*-value are shown (with WT genotype as reference group). Furthermore, the two most frequently occurring CPA-based grouped annotations are depicted. If this top 2 contained shared positions ending up with >3 processes, only the main occurring grouped annotation was shown (indicated with *)

module annotations, blood T2 modules demonstrated several annotations that were not found in blood T1, such as "histone modification," "chromosome organization," "ribosome biogenesis," "tumor suppressor activity," and "cholesterol" (Figs. 6, 7, 8) indicating disease progression.

Gene Expression Changes Compared Between Brain and Blood

Differential Gene Expression Analysis Shows a Small Overlap Between Brain and Blood

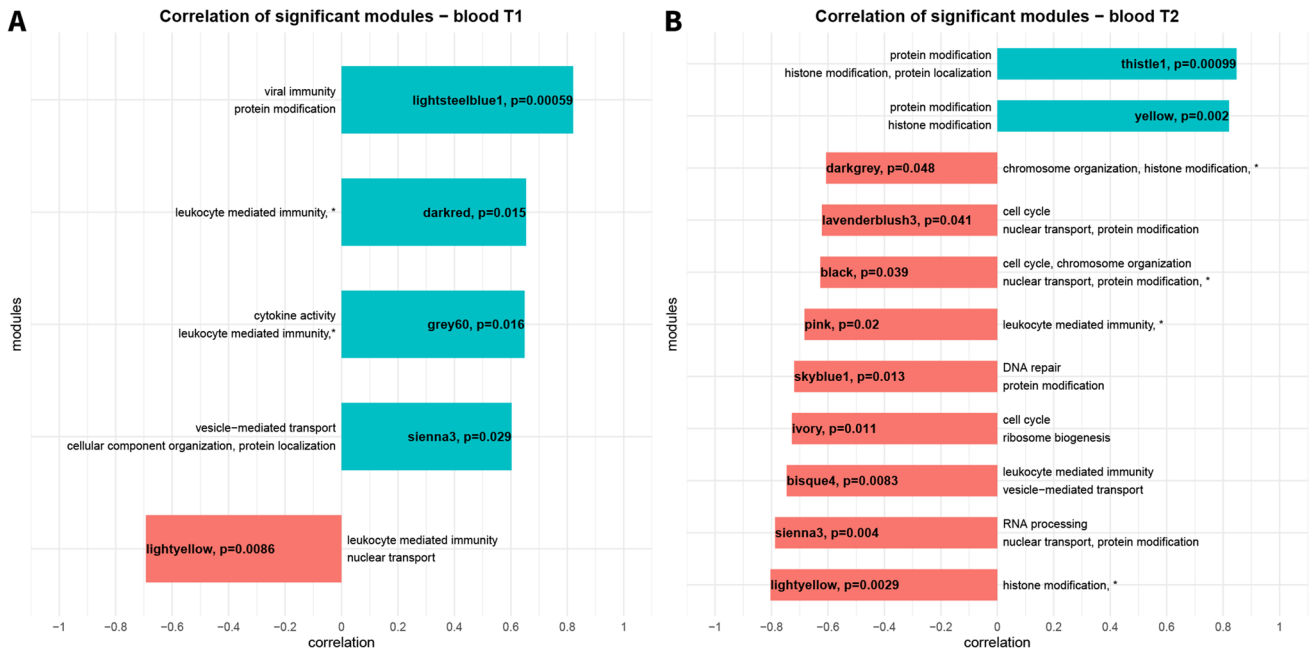


Fig. 5 Overview of significant WGCNA modules in blood at time point T1 (A) and T2 (B). The correlation and *P*-value are shown for each module (with WT genotype as reference group). Furthermore, the two most frequently occurring CPA-based grouped annotations are depicted. If this top 2 contained shared positions ending up with >3 processes, only the main occurring grouped annotation was shown (indicated with *)

After confirming HD pathology in YAC128 mice, we wanted to assess whether gene expression changes observed in brain are reflected in blood. First, we compared the results of differential gene expression analysis in brain and blood described above. The 30 DE genes with highest absolute fold change for each tissue are combined in Fig. 6 and their differential expression was assessed in all regions. Only two of these genes were differentially expressed in both blood and brain (*Cmpk2* and *Nt5e*), confirming poor preservation of gene expression changes among these two tissues in line with previous studies [10]. Nevertheless, 16 genes were differentially expressed in multiple brain regions, among which 3 genes in all four regions (*Zfp488*, *Acy3*, and *Il33*). Considering the blood tissue, 4 DE genes occurred in both time points (*Cmpk2*, *Rtp4*, *Oas2*, *Ephb6*).

WGCNA-Based Analysis Reveals Overlap in Brain and Blood Pathology

Since the limited overlap between brain and blood on gene level, we compared HD-associated gene expression changes on a functional level using the grouped annotations of the WGCNA modules. To examine the disease-associated processes shared between brain and blood, we studied the overlap between the grouped module annotations (Figs. 7, 8, Supplementary Table 5). In total, 23 grouped module annotations overlapped between brain and both time points in blood, such as "apoptosis," "nuclear transport," "protein localization," and "vesicle-mediated transport." Three grouped module annotations were shared between brain and only blood T1 specifically: "intracellular transport," "cell recognition," and "cell adhesion." Seven grouped module annotations were shared between brain and blood time point T2. Among these were "cholesterol," "ribosome biogenesis," and "histone modification."

Blood-Brain Overlap in YAC128 Mouse Model Reflects Findings in Human HD

We compared the disease-specific signatures identified in the YAC128 mice with the previously identified signatures in human HD brain and blood [8] to assess how well the YAC128 model reflects HD pathology in blood. The original annotations from these human HD-associated modules from three brain regions (caudate nucleus, Brodmann area 4, cerebellum) and blood identified by Mina et al. [8] were grouped similarly as was done with the mouse modules, ending up with 38 unique grouped annotations that showed overlap between human brain and blood. From the 33 annotations that were found both in mouse blood and brain, 24 were also found in the overlap of annotations between human blood and brain (Fig. 9, Supplementary Table 6). These 24 annotations included the human blood-brain disease-specific processes highlighted by Mina et al.

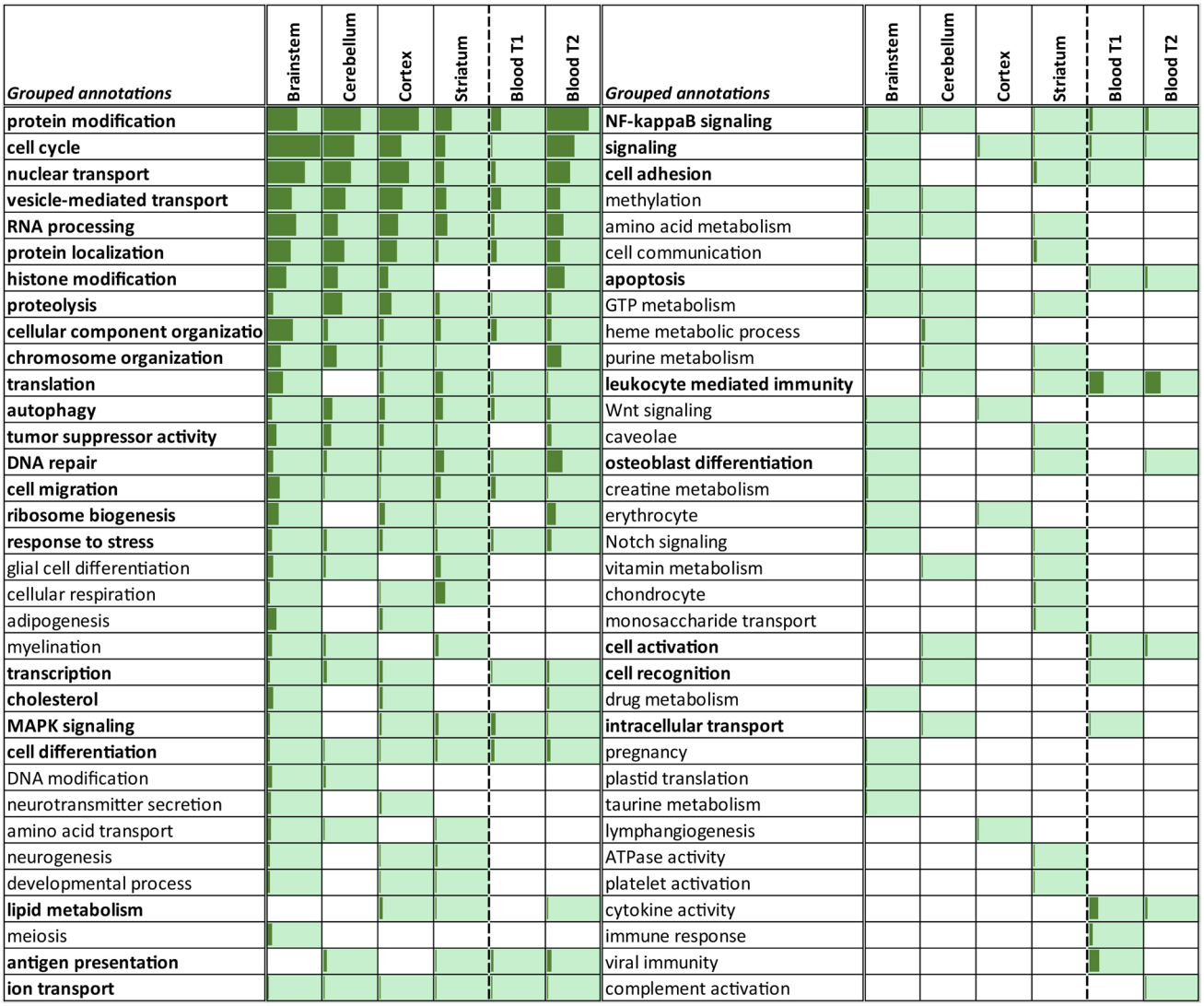


Fig. 6 Occurrence of grouped annotations for HD-associated modules in brain regions and blood time points. The presence of a grouped annotation in one or more HD-associated modules of a brain region or blood time point 1 (T1) or 2 (T2) is indicated by the light green color while the dark green bar depicts how often the grouped annotation occurred in the HD-associated modules of this region (full bar = 44 counts). Bold annotations occurred both in brain and blood

[8], such as "leukocyte-mediated immunity," "cell cycle," "protein modification," and "protein transport," suggesting that the YAC128 model reflects human disease to a large extent. However, the annotations "immune response" and "cytokine activity" were shared in human blood and brain but were not present in mouse brain, suggesting that the immune component is less prominent in mice.

Blood Pathology Showing Similarity with Brain Pathology

We assessed the similarity between brain and blood modules (in terms of the total number of overlapping module annotations) in order to prioritize blood modules that exhibit a similar pathology to brain and could be potentially used for monitoring the disease. Considering that overlap will always be found to some extent, we assessed the statistical significance of the pairwise similarity of the HD-associated blood modules with each brain module (both associated with HD and not). To that end, we performed a randomization experiment in which the overlap in the original annotations between two modules was compared to the pairwise overlapping annotations between one hundred random modules of equal size generated for both original modules from each network. The *P*-value obtained by this permutation test indicates the specificity of the original overlap found when compared to what could be expected with random annotations. The module pairs were ranked based on the *P*-value to indicate

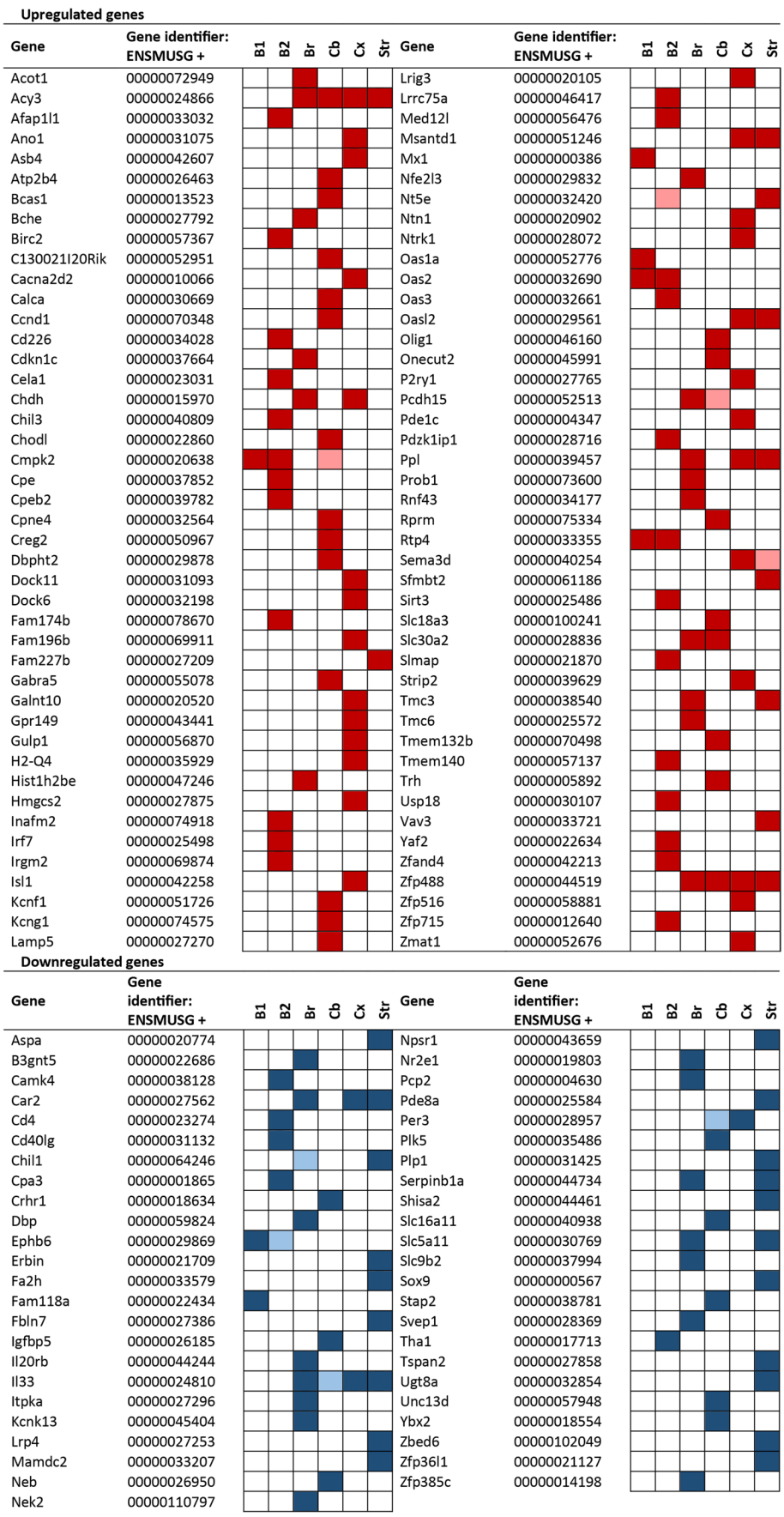


Fig. 7 Results of DGE analysis shown for the top 30 differentially expressed genes of each tissue. Top 30 differentially expressed genes based on absolute fold change for each tissue, either being up- (red), down-regulated (blue) or not differentially expressed (white). When differentially expressed but not present in the top 30 of this tissue, the gene is depicted either light red or light blue. B1: blood time point 1; B2: blood time point 2; Br: brainstem; Cb: cerebellum; Cx: cortex; Str: striatum

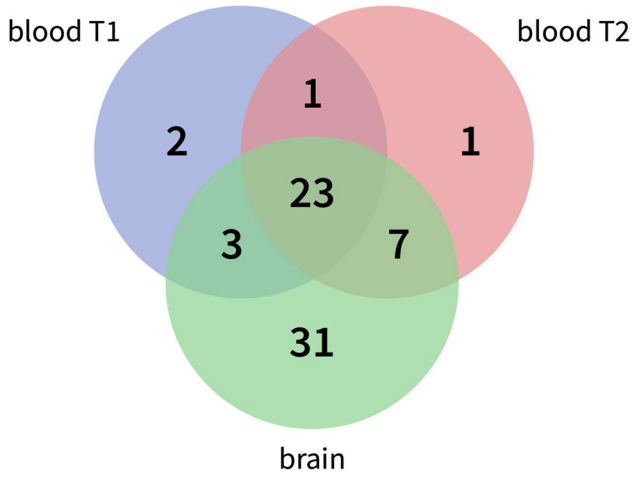


Fig. 8 Overlap between the grouped annotations assigned to HD-associated modules in blood time points and brain. Comparison shown of grouped annotations from the first (blood T1) and second (blood T2) time point in blood and the four brain regions investigated combined (brain: brainstem, cerebellum, cortex, striatum)

the specificity of an observed overlap (Supplementary File IV). Multiple testing correction to the *P*-values of the pairwise overlap between two regions yielded no significant overlap (FDR < 0.05, data not shown). Nonetheless, using the ranking based on the unadjusted *P*-value, we could identify HD-associated blood modules that rather showed overlap with HD-associated brain modules than with non-HD associated brain modules, indicating similar pathology. Assuming that functional overlap of blood modules with HD pathology in brain helps to identify HD-specific processes in blood, this group of blood modules could be valuable in monitoring HD in blood.

For further investigation regarding the use of the identified blood modules for disease monitoring, we selected the blood modules with an absolute correlation with the HD genotype >0.65. Among these were two blood T1 (*lightyellow*, *sienna3*) and seven blood T2 modules (*bisque4*, *lightyellow*, *pink*, *sienna3*, *skyblue1*, *thistle1*, and *yellow*) that demonstrated predominant overlap with HD pathology in brain. To select representative genes for each blood module, we performed connectivity analysis to identify hub genes as these are hypothesized to have a central regulatory function within the module. For each module, the top 10% of genes with the highest intramodular connectivity were defined as hub genes. In addition to their intramodular connectivity, we looked at the unadjusted *P*-value of differential expression for these genes, to see whether they would be able to distinguish disease from control. As gene expression must be sufficiently high to be detected, we also assessed the log counts per million (logCPM) indicating the expression level (Fig. 10A).

Based on the range of the differential expression fold change, *P*-value, and counts per million of the hub genes, the most promising modules presenting pathology shared with brain were blood T1 modules *lightyellow* and blood T2 modules *pink*, *sienna3*, *yellow*, and *thistle1*. Blood T1 module *lightyellow* represented T cell immunity and nuclear transport and showed predominant overlap with HD-associated brain modules. In particular, it resembled cerebellar modules *mediumpurple3*, showing antigen presentation and lymphocyte activation, and *blue4*, involved in nuclear transport. All hub genes detected were differentially expressed (unadjusted *P*-value <0.05), of which the following remained significant after multiple testing correction: *Fam118a* (log fold change (logFC) = −0.56), *Ephb6* (logFC = −0.65). Blood T2 module *pink* also showed overlap with cerebellar module *mediumpurple3*. Hub gene *Tcf7*, mainly expressed by T cells, was differentially expressed, although did not reach significance after multiple testing correction (logFC = −0.69). Blood T2 module *sienna3* was related to RNA splicing and nuclear transport: annotations that were shared with brainstem modules *salmon4*, *lightcyan1*, and *darkmagenta* and cerebellar module *blue4*. Hub genes showing the strongest downregulation were *Orm2* (logFC = −0.64; *P*-value < 0.05; FDR < 0.05) and *Dgka* (logFC = −0.75; *P*-value < 0.05; FDR > 0.05). Blood T2 module *yellow* overlapped with cerebellar modules *firebrick*, *darkorange*, and *coral1*, related to protein and histone modification. Several hub genes were differentially expressed and showed high fold change, for instance *Med12l* (logFC = 1.45) and *Tmem140* (logFC = 1.11). Blood T2 module *thistle1* showed the highest positive correlation with the HD genotype and depicted overlap with the cerebellar modules *darkorange* and *coral1*, based on annotations related to protein modification. Among the hub genes that were differentially expressed were the genes *Stx11* (logFC = 0.69), *Mbd2* (logFC = 0.67), *Ankib1* (logFC = 0.54), and *Sdcbp* (logFC = 0.52).

In addition to the annotation-based analysis, we performed module preservation analysis to identify HD-associated blood modules that are preserved in brain, based on network properties [22]. This method evaluates whether a module in the blood network (reference network) can be found in the brain networks (test network) by comparing the connectivity and density of the genes in the blood module to the test network. Evidence for preservation is weak when a *Z*-score between 2 and 10 is obtained and strong for Z-scores above 10. We found a few blood modules in preserved brain networks, with weak preservation scores (Supplementary Fig. 10). This is in line with previous research where gene preservation between blood and brain in control samples was found to be poor [10]. Two HD-associated blood T2 modules showed weak preservation in striatum (*black* and *lavenderblush3*) and

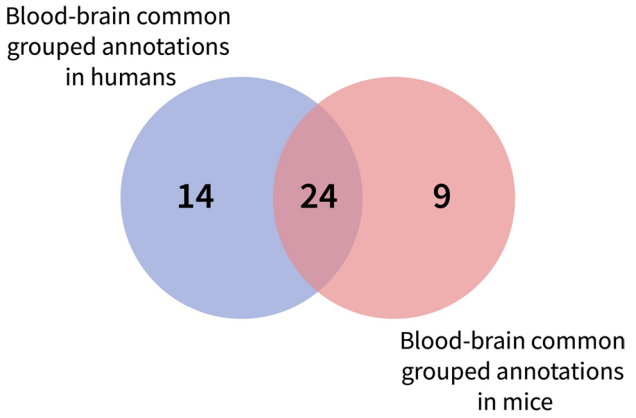


Fig. 9 Overlap of blood-brain common grouped annotations between YAC128 mice and HD patients. In mice, 33 grouped annotations were shared between blood and brain. Of these, 24 occurred in the human common grouped annotations as well

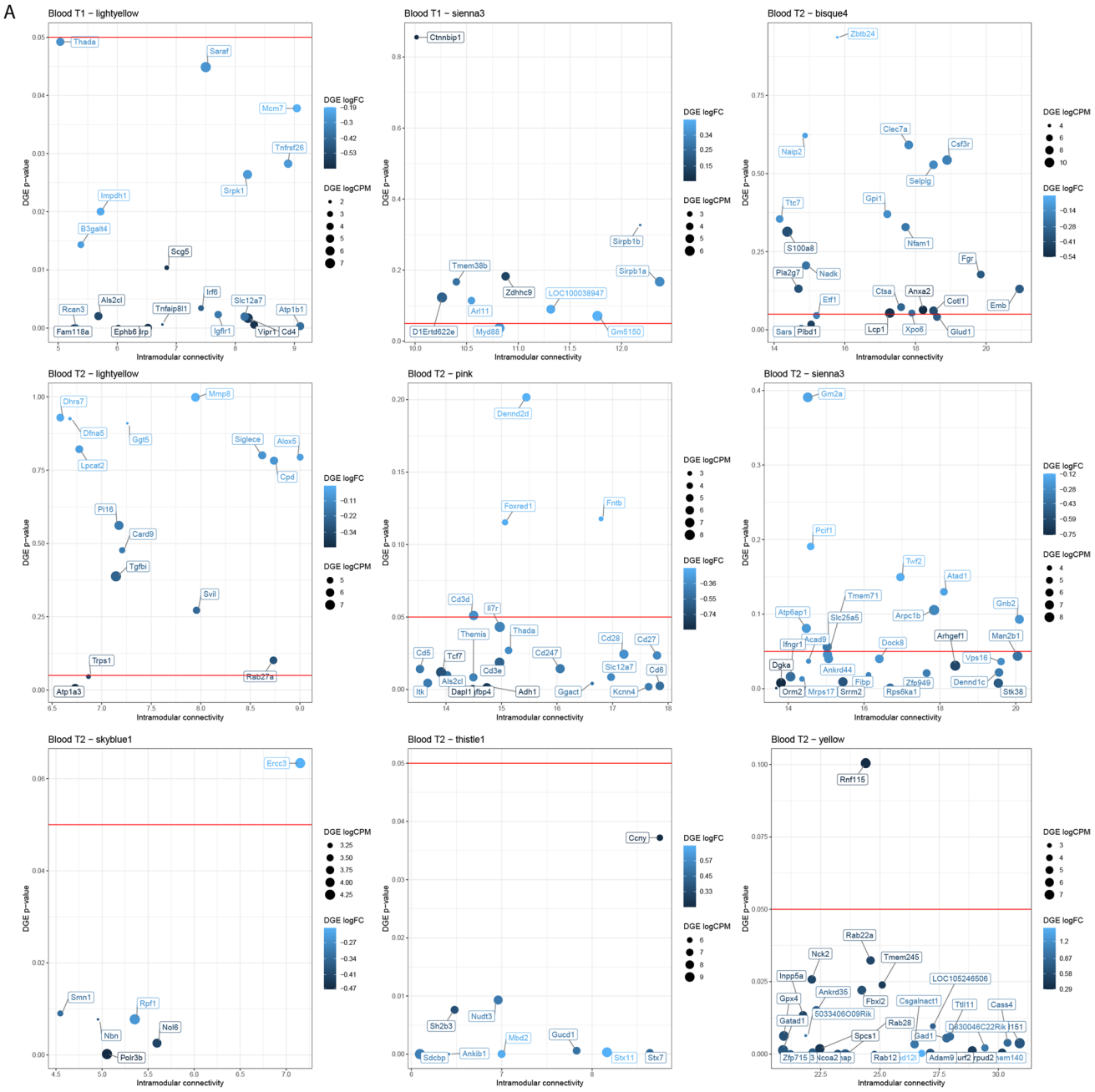
the HD-associated blood T2 module *yellow* was preserved in brainstem and striatum. There was a minor overlap between the functional annotation-based analysis and the module preservation analysis, exhibiting again their complementary character as was found in Mina et al. [8]. The only overlap between the two approaches was the blood T2 module *yellow*, as the other two identified by the preservation analysis (*black* and *lavenderblush3*) were not passing our criteria for HD association (presenting with $r<|0.65|$).

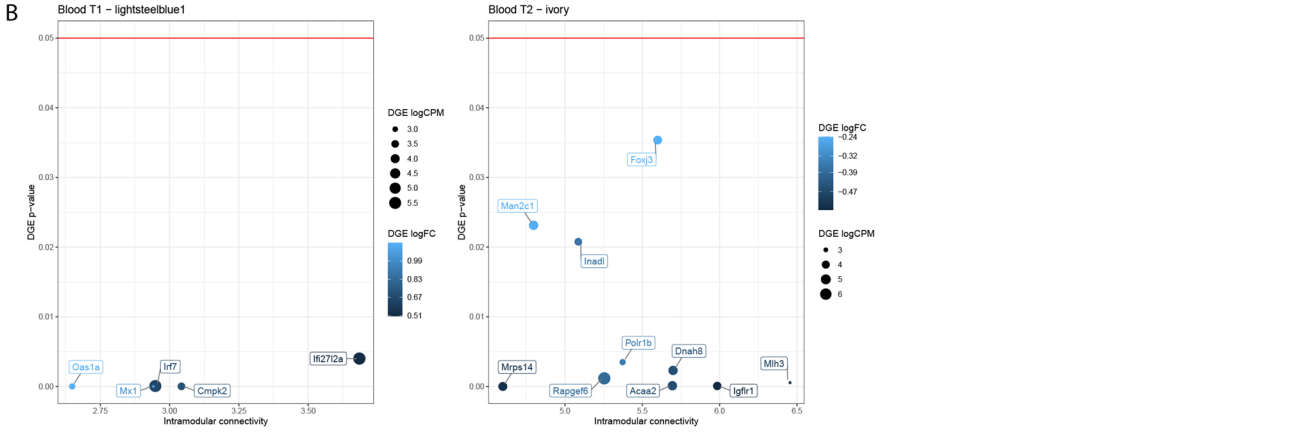
Blood Modules Presenting Blood-Specific Pathology

Among the modules in blood that were highly associated with HD ($r > |0.65|$), there were two modules that exhibited a blood specific pathology as they did not show predominant overlap with HD-associated modules in brain (Supplementary File IV). Blood T1 module *lightsteelblue1* was highly correlated to HD ($r = 0.82$) and was mainly related to viral immune response, but also to cell migration. The hub genes included the differentially expressed genes *Oas1a* (logFC = 1.15), *Mx1* (logFC = 1.01), and *Cmpk2* (logFC = 0.71) (Fig. 10B). Furthermore, *lightsteelblue1* presented *Irf7* as hub gene though not significant after multiple testing correction. Regarding blood T2, the module ivory exhibited a similar behavior. This module was related to ribosome assembly, DNA repair, and cell cycle check and presented hub genes that showed differential expression, such as *Igflr1* (logFC = −0.55), *Mlh3* (logFC = −0.50), *Acaa2* (logFC = −0.46), and *Mrps14* (logFC = −0.53). Hub gene *Dnah8* (logFC = −0.46) seemed to be differentially expressed as well (P-value < 0.01, FDR = 0.08).

Longitudinal Hub Gene Expression Differs Between YAC128 and Control Mice

Due to their high amount of connectivity within a module, hub genes are hypothesized to reflect key drivers of change and, thus, they could serve as potential biomarkers. We sought to assess whether the hub genes identified in T1 and T2 could serve as biomarkers indicating disease progression. We studied the expression changes from T1 to T2 for the aforementioned hub genes (*Fam118a*, *Ephb6*, *Tcf7*, *Orm2*, *Dgka*, *Med12l*, *Tmem140*, *Stx11*, *Mbd2*, *Ankib2*, *Sdcbp*, *Oas1a*, *Mx1*, *Cmpk2*, *Irf7*, *Igflr1*, *Mlh3*, *Acaa2*, *Mrps14*, *Dnah8*) (Fig. 11, Supplementary Fig. 11, Supplementary Table 7). For genes showing a significant difference in change over time between WT and YAC128 mice, longitudinal measurement could be informative regarding disease progression. The change in expression from T1 to T2 was significantly different between WT and YAC128 mice for the following genes: *Acaa2*, *Ankib1*, *Mbd2*, *Mrps14*, *Orm2*, *Sdcbp*, *Stx11* (Fig. 11). *Acaa2* and *Mrps14* were part of T2 *ivory* module (reflecting cell cycle and ribosome biogenesis). Both are mitochondrial proteins and show a tendency towards a negative slope in YAC128 mice in contrast to wild-type mice. *Acaa2*, acetyl-CoA acyltransferase 2, is involved in the final step of the mitochondrial fatty acid beta-oxidation spiral. *Mrps14*, mitochondrial ribosomal protein S14, is involved in mitochondrial translation. *Sdcbp*, *Mbd2*, *Ankib1*, and *Stx11* are hub genes of T2 *thistle1* (representing protein modification, histone modification, protein localization). Whereas wild-type mice show downregulation of these genes over time, YAC128 mice exhibit increasing levels of expression over time. *Sdcbp*, syndecan-binding protein, is an adapter protein involved in for instance trafficking of transmembrane proteins. *Mbd2*, methyl-CpG-binding domain protein 2, binds methylated CpG islands in promoters and recruits histone deacetylases and DNA methyltransferases, acting as transcriptional repressor. *Ankib1*, ankyrin repeat and IBR domain containing 1, is involved in ubiquitination, being part of E3 ligase complex. *Stx11*, syntaxin 11, is a SNARE protein regulating protein trafficking from late endosomes to the trans-Golgi network. *Orm2*, orosomucoid 2, hub gene of T2 *sienna3* (representing RNA processing, nuclear transport, and protein modification), is a transport protein thought to be involved in immune activity during an acute-phase reaction. Whereas levels are stable to increasing in wild-type mice, YAC128 mice show decreasing levels over time.


A
Blood T1 – lightyellow
Blood T1 – sienna3
Blood T2 – bisque4
Blood T2 – lightyellow
Blood T2 – pink
Blood T2 – sienna3
Blood T2 – skyblue1
Blood T2 – thistle1
Blood T2 – yellow
DGE p-value
Intramodular connectivity
DGE logFC
DGE logCPM


B
Blood T1 – lightsteelblue1
Blood T2 – ivory
DGE p-value
Intramodular connectivity
DGE logFC
DGE logCPM

◂**Fig. 10** Hub genes and their information from differential gene expression analysis. **A** HD-associated (r>0.65) blood modules showing predominant overlap with HD-associated modules in at least one brain region; **B** HD-associated (r>0.65) blood modules that did not predominantly show overlap with HD-associated brain modules suggesting blood-specific pathology

Discussion

In this study, we identified HD disease signatures in RNAseq data from brain and blood of the YAC128 mouse model using WGCNA. After validation of our RNAseq results by qPCR and western blot, we used WGCNA to identify 5 to 17 clusters of co-expressed genes (modules) per tissue associated with the HD genotype. There was a large overlap in the disease-related processes between brain and blood, which was highly comparable to previous findings in human HD brain and blood both showing for instance cell cycle, leukocyte-mediated immunity, and protein modification [8].

The HD-associated modules we identified in YAC128 brain exhibited a wide range of deregulated processes in accordance to previous studies on the molecular pathogenesis of human HD and various mouse models [2, 28–31], such as altered protein modification, cell cycle, cellular transport, RNA processing, autophagy, proteasome degradation, chromatin organization, and cellular respiration. We also detected processes specific for the nervous system previously described in HD such as synaptic function, myelination, and glial cell differentiation [28, 31, 32]. However, the highest number of HD-associated modules was identified in brainstem and cerebellum, followed by cortex and striatum, whereas Mina et al. identified most changes in the human HD brain in striatum [8]. These discrepancies are to be expected as the human brain tissue reflects the end stage of the disease, showing vast striatal pathology, and the YAC128 mouse model shows a modest phenotype reflecting earlier stages of the disease. Interestingly, both human and murine data identify more HD-associated modules in cerebellum compared to cortex which is in line with increasing evidence of cerebellar pathology in human HD brain [33].

We observed a progression of the disease in blood, indicated by a higher amount of differentially expressed genes and more HD-associated modules at blood time point 2 (T2) compared to time point 1 (T1). HD-associated modules at T1 were mainly related to the immune response, HD-associated modules at T2 were mostly related to processes such as cholesterol, cell cycle, protein transport, chromosome organization, and DNA repair. The progression in cholesterol disturbance has been found before in YAC128 blood plasma by Valenza et al. [34], who showed decreased plasma levels of cholesterol and its precursors.

As we compared YAC128 brain and blood, we detected a large overlap of 33 grouped module annotations within their HD-associated modules, including protein modification, nuclear transport, and RNA splicing. Of these 33 shared grouped annotations, 23 were found in both time points in blood, but seven grouped module annotations were only associated with the latest time point in blood including ribosome biogenesis and cholesterol. This indicates that the wide range of deregulation induced by mutant HTT seen in brain becomes more apparent in blood as disease progresses.

The relevance of our findings to the pathology found in patients was indicated by the large overlap with previous results in human HD [8]. Of the 33 common disease–associated grouped annotations in YAC128, 24 were identified as common blood-brain processes in human HD as well by Mina et al. including cell cycle, protein modification, and protein transport [8]. A study comparing human blood and brain from HD patients using a different approach also found deregulation of RNA splicing and processing and DNA repair [33]. However, these studies comparing human blood and brain both highlight immune processes as shared disease processes [8, 35]. In contrast, although we find immune-related processes in HD-associated modules, these are not as prominent as the immune response shown in brain tissue from human HD [8, 35, 36]. This difference is in line with the poor preservation of human immune-related modules in brains of HD mouse models (including YAC128) found by Neueder et al. [34] and could be due to differences between human and mouse immunology [37].

The biomarker potential of the processes we identified is confirmed by several other studies. For instance, Puorro and colleagues showed an increase in autophagy marker expression in peripheral blood mononuclear cells of HD patients using qPCR [38]. Areal et al. detected downregulation of the microtubule-associated motor protein dynein axonemal heavy chain 6 (DNAH6), involved in cellular component organization, by microarray analysis in striatum of hQ111/Q111 mice and confirmed this with qPCR in both mice striatum and blood of HD patients [39]. In addition, Castaldo et al. recently described the potential of assessing telomere length with qPCR and DNA double-strand breaks using fluorescence-activated cell sorting (FACS) as predictive markers [40]. As we also described autophagy, cellular component organization, and DNA repair, our study provides more evidence for aforementioned markers. Interestingly, we identified the dynein axonemal heavy chain 8 (DNAH8) as hub gene of HD-associated blood T2 module *ivory*.

Furthermore, our study identified modules that showed blood-specific pathology. These modules were highly correlated with the HD genotype but showed no predominant overlap with brain pathology. Blood-specific pathology at T1 was related to viral immunity and cell migration, while the module annotations at T2 pointed at ribosome assembly, DNA repair, and cell cycle.

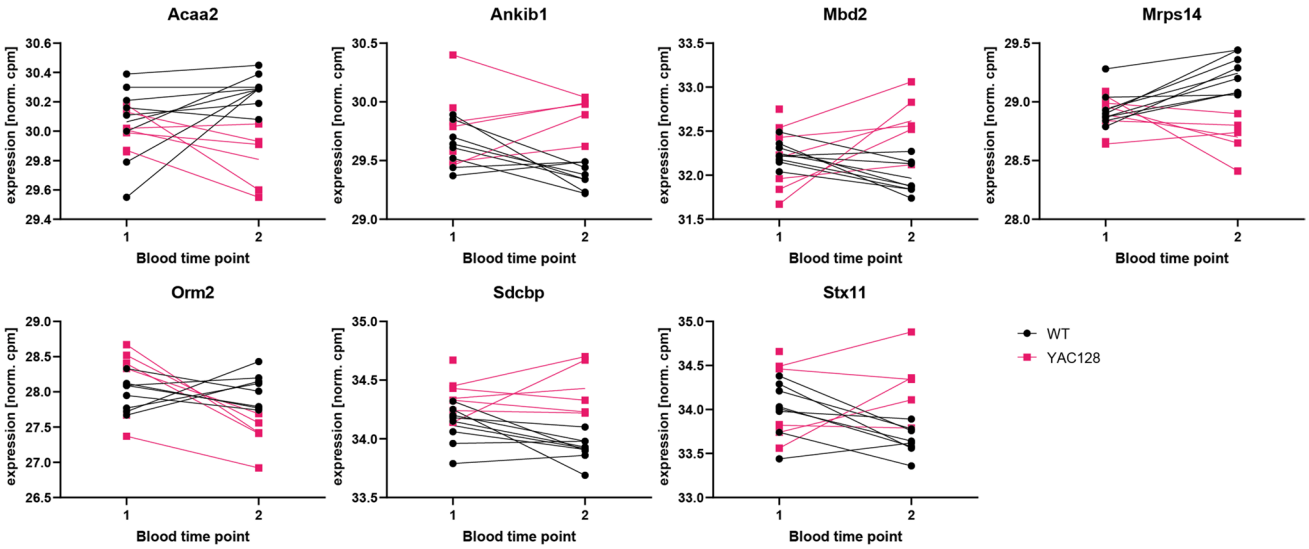


Fig. 11 Longitudinal hub gene expression for YAC128 and control mice. Expression levels as measured by RNAseq on T1 (1) and T2 (2) for wild-type (WT, black) and YAC128 mice (pink). The slopes for wild-type and YAC128 mice were significantly different ($P < 0.05$, see Suppl. Table 7). (n=6–8 mice per group, norm. cpm: normalized counts per million)

Finally, our findings provide additional gene targets that could be used as biomarkers. It is generally thought that a complementary set of biomarkers will be required to successfully monitor and predict HD [7]. This is also emphasized by Mastrokolias et al., who previously proposed a blood biomarker gene panel by correlating gene expression with clinical scales, including five genes involved in different functions, such as inflammation and circadian rhythm [41]. The authors emphasize that different sets of biomarkers should be developed and validated, involving multiple genes [41]. Moreover, our analysis at the functional level argues for a systems-level approach. Rather than focussing on the behavior of individual genes, even in a panel, the combined behavior of a group of functionally related genes may prove a more robust biomarker for HD. The hub genes we present that showed expression levels sufficient for detection by qPCR and differential expression between HD and control would be suitable for further development of a monitoring panel. Interestingly, 3 of the 5 hub genes (*Irf7*, *Mx1*, *Cmpk2*) that we identified in blood T1 module *lightsteelblue1* were found to be differentially expressed in human peripheral blood as well in a recent RNAseq study [42].

From the identified hub genes, 7 genes (*Acaa2*, *Ankib1*, *Mbd2*, *Mrps14*, *Orm2*, *Sdcbp*, *Stx11*) showed a significant change in expression from T1 to T2 between wild-type and YAC128 mice. These genes are involved in HD-related processes we identified, such as vesicle-mediated transport (*Stx11*), protein localization (*Sdcbp*), protein modification (*Ankib1*), regulation of transcription (*Mbd2*), and immunity (*Orm2*). Interestingly, *Acaa2* and *Mrps14* are both mitochondrial proteins, corresponding to the mitochondrial dysfunction generally implicated in HD [28]. To assess their potential application as biomarkers of disease progression, future research into their longitudinal expression during the life of YAC128 mice is required. In addition to that, their correlation with other phenotypic markers or response to treatment will need to be assessed.

Despite the promising nature of the disease signatures, we identified for monitoring disease pathology in HD, our study had however some limitations. The sample size within our data sets ranged between 12 and 14 mice, which is just below the recommended number of 15 for WGCNA [21]. The lower sample size increased the vulnerability to outlying data. We tried to reduce this by removing outliers after sample outlier analysis. Secondly, for the module annotation, we applied the same literature database that was used by Mina [8], as we wanted to compare the murine results with the human findings that were discovered in the same publication. However, this database has literature information up to March 2012 and certain specific functional annotations might change for both human and mouse.

Conclusions

We observed progression in blood pathology in the YAC128 HD mouse model, showing similarities with disease-related processes that occur in brain but also identifying HD pathology that was specific for blood. Disease signatures shared between mouse blood and brain show high overlap with signatures present in human blood and brain. Using network

and expression characteristics, we selected hub genes as representatives for the identified disease signatures. Our systems approach provides several biological processes and genes for further investigation of their role in monitoring disease in blood.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12035-021-02680-8.

Acknowledgements The authors want to thank Melvin M. Evers for his help in the design of the animal studies.

Author Contribution Wet lab experiments were performed by M. O., E. C. K., and L. J. A. T. Analysis of RNA sequencing was done by E. C. K. and E. M. R. T. gave advice on statistical analysis. Experiments were designed by L. J. A. T., M. R., W. v. R. M., E. M., and K. H. Interpretation of data by E. C. K., M. R., W. v. R. M., and E. M. Writing of the paper was done by E. C. K. All authors read and approved the final manuscript.

Funding This work was partly funded by Campagne Team Huntington. Sequencing was funded by the European Union seventh Framework Program (FP7/2007-2013), grant agreement no. 305,121 (Neuromics).

Availability of Data and Materials The datasets generated and analyzed during the current study are available in the GEO repository, GSE170998.

Code Availability The code used for the analyses is available at Zenodo: https://zenodo.org/record/5806116.

Declarations

Ethics Approval All animal experiments were licensed by the Central Committee for Animal experiments (CCD) in AVD1160020171069, valid from 1 September 2017 until 1 September 2022.

Consent to Participate Not applicable.

Consent for Publication Not applicable.

Competing Interests Not applicable.

References

1. Nopoulos PC (2016) Huntington disease: a single-gene degenerative disorder of the striatum. Dialogues Clin Neurosci 18(1):91–98
2. Bano D, Zanetti F, Mende Y, Nicotera P (2011) Neurodegenerative processes in Huntington's disease. Cell Death Dis 2:e228. https://doi.org/10.1038/cddis.2011.112
3. Carroll JB, Bates GP, Steffan J, Saft C, Tabrizi SJ (2015) Treating the whole body in Huntington's disease. Lancet Neurol 14(11):1135–1142. https://doi.org/10.1016/S1474-4422(15)00177-5
4. The Huntington's Disease Collaborative Research Group (1993) A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. Cell 72 (6):971–983 https://doi.org/10.1016/0092-8674(93)90585-e
5. Rodrigues FB, Wild EJ (2020) Huntington's Disease Clinical Trials Corner: April 2020. J Huntingt Dis Preprint 1–13 https://doi.org/10.3233/JHD-200002
6. Shannon KM, Fraint A (2015) Therapeutic advances in Huntington's Disease. Mov Disord 30(11):1539–1546. https://doi.org/10.1002/mds.26331
7. Zeun P, Scahill RI, Tabrizi SJ, Wild EJ (2019) Fluid and imaging biomarkers for Huntington's disease. Mol Cell Neurosci. https://doi.org/10.1016/j.mcn.2019.02.004
8. Mina E, van Roon-Mom W, Hettne K, van Zwet E, Goeman J, Neri C, P ACTH, Mons B, Roos M (2016) Common disease signatures from gene expression analysis in Huntington's disease human blood and brain. Orphanet J Rare Dis 11(1):97. https://doi.org/10.1186/s13023-016-0475-2
9. Zadel M, Maver A, Kovanda A, Peterlin B (2018) Transcriptomic biomarkers for Huntington's Disease: are gene expression signatures in whole blood reliable biomarkers? Omics 22(4):283–294. https://doi.org/10.1089/omi.2017.0206
10. Cai C, Langfelder P, Fuller TF, Oldham MC, Luo R, van den Berg LH, Ophoff RA, Horvath S (2010) Is human blood a good surrogate for brain tissue in transcriptional studies? BMC Genomics 11:589. https://doi.org/10.1186/1471-2164-11-589
11. Slow EJ, van Raamsdonk J, Rogers D, Coleman SH, Graham RK, Deng Y, Oh R, Bissada N, Hossain SM, Yang YZ, Li XJ, Simpson EM, Gutekunst CA, Leavitt BR, Hayden MR (2003) Selective striatal neuronal loss in a YAC128 mouse model of Huntington disease. Hum Mol Genet 12(13):1555–1567
12. Toonen LJA, Overzier M, Evers MM, Leon LG, van der Zeeuw SAJ, Mei H, Kielbasa SM, Goeman JJ, Hettne KM, Magnusson OT, Poirel M, Seyer A, t Hoen PAC, van Roon-Mom WMC (2018) Transcriptional profiling and biomarker identification reveal tissue specific effects of expanded ataxin-3 in a spinocerebellar ataxia type 3 mouse model. Mol Neurodegener 13(1):31https://doi.org/10.1186/s13024-018-0261-9
13. Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26(1):139–140. https://doi.org/10.1093/bioinformatics/btp616
14. McCarthy DJ, Chen Y, Smyth GK (2012) Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucleic Acids Res 40(10):4288–4297. https://doi.org/10.1093/nar/gks042
15. Hansen KD, Irizarry RA, Wu Z (2012) Removing technical variability in RNA-seq data using conditional quantile normalization. Biostatistics 13(2):204–216. https://doi.org/10.1093/biostatistics/kxr054
16. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate - a practical and powerful approach to multiple testing. J R Stat Soc B Methodol 57(1):289–300
17. Untergasser A, Nijveen H, Rao X, Bisseling T, Geurts R, Leunissen JA (2007) Primer3Plus, an enhanced web interface to Primer3. Nucleic Acids Res 35 (Web Server issue):W71–74 https://doi.org/10.1093/nar/gkm306
18. Ramakers C, Ruijter JM, Deprez RH, Moorman AF (2003) Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. Neurosci Lett 339(1):62–66
19. Benjamini Y, Krieger AM, Yekutieli D (2006) Adaptive linear step-up procedures that control the false discovery rate. Biometrika 93(3):491–507. https://doi.org/10.1093/biomet/93.3.491
20. Oldham MC, Langfelder P, Horvath S (2012) Network methods for describing sample relationships in genomic datasets: application to Huntington's disease. BMC Syst Biol 6:63. https://doi.org/10.1186/1752-0509-6-63
21. Langfelder P, Horvath S (2008) WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 9:559. https://doi.org/10.1186/1471-2105-9-559

22. Langfelder P, Luo R, Oldham MC, Horvath S (2011) Is my network module preserved and reproducible? PLoS Comput Biol 7(1):e1001057. https://doi.org/10.1371/journal.pcbi.1001057
23. Jelier R, Schuemie MJ, Roes PJ, van Mulligen EM, Kors JA (2008) Literature-based concept profiles for gene annotation: the issue of weighting. Int J Med Inform 77(5):354–362. https://doi.org/10.1016/j.ijmedinf.2007.07.004
24. Jelier R, Schuemie MJ, Veldhoven A, Dorssers LC, Jenster G, Kors JA (2008) Anni 2.0: a multipurpose text-mining tool for the life sciences. Genome Biol 9(6):R96 https://doi.org/10.1186/gb-2008-9-6-r96
25. University P Generic GO Term Mapper. https://go.princeton.edu/cgi-bin/GOTermMapper. Accessed 10 Nov 2021
26. VIB/UGent Calculate and draw custum Venn diagrams. http://bioinformatics.psb.ugent.be/webtools/Venn/. Accessed 3 Aug 2020
27. Westfall PH, Young SS (1993) Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment. John Wiley, New York
28. Jimenez-Sanchez M, Licitra F, Underwood BR, Rubinsztein DC (2017) Huntington's disease: mechanisms of pathogenesis and therapeutic strategies. Cold Spring Harb Perspect Med 7 (7) https://doi.org/10.1101/cshperspect.a024240
29. Skotte NH, Andersen JV, Santos A, Aldana BI, Willert CW, Nørremølle A, Waagepetersen HS, Nielsen ML (2018) Integrative characterization of the R6/2 mouse model of Huntington's disease reveals dysfunctional astrocyte metabolism. Cell Rep 23(7):2211–2224. https://doi.org/10.1016/j.celrep.2018.04.052
30. Bayram-Weston Z, Stone TC, Giles P, Elliston L, Janghra N, Higgs GV, Holmans PA, Dunnett SB, Brooks SP, Jones L (2015) Similar striatal gene expression profiles in the striatum of the YAC128 and HdhQ150 mouse models of Huntington's disease are not reflected in mutant Huntingtin inclusion prevalence. BMC Genomics 16(1):1079. https://doi.org/10.1186/s12864-015-2251-4
31. Teo RT, Hong X, Yu-Taeger L, Huang Y, Tan LJ, Xie Y, To XV, Guo L, Rajendran R, Novati A, Calaminus C, Riess O, Hayden MR, Nguyen HP, Chuang KH, Pouladi MA (2016) Structural and molecular myelination deficits occur prior to neuronal loss in the YAC128 and BACHD models of Huntington disease. Hum Mol Genet 25(13):2621–2632. https://doi.org/10.1093/hmg/ddw122
32. Wilton DK, Stevens B (2020) The contribution of glial cells to Huntington's disease pathogenesis. Neurobiol Dis 143:104963. https://doi.org/10.1016/j.nbd.2020.104963
33. Franklin GL, Camargo CHF, Meira AT, Lima NSC, Teive HAG (2021) The role of the cerebellum in Huntington's disease: a systematic review. Cerebellum (London, England) 20(2):254–265. https://doi.org/10.1007/s12311-020-01198-4
34. Valenza M, Carroll JB, Leoni V, Bertram LN, Bjorkhem I, Singaraja RR, Di Donato S, Lutjohann D, Hayden MR, Cattaneo E (2007) Cholesterol biosynthesis pathway is disturbed in YAC128 mice and is modulated by huntingtin mutation. Hum Mol Genet 16(18):2187–2198. https://doi.org/10.1093/hmg/ddm170
35. Hensman Moss DJ, Flower MD, Lo KK, Miller JRC, van Ommen G-JB, 't Hoen PAC, Stone TC, Guinee A, Langbehn DR, Jones L, Plagnol V, van Roon-Mom WMC, Holmans P, Tabrizi SJ (2017) Huntington's disease blood and brain show a common gene expression pattern and share an immune signature with Alzheimer's disease. Sci Rep 7:44849 https://doi.org/10.1038/srep44849. https://www.nature.com/articles/srep44849#supplementary-information
36. Neueder A, Bates GP (2014) A common gene expression signature in Huntington's disease patient brain regions. BMC Med Genomicshttps://doi.org/10.1186/s12920-014-0060-2
37. Mestas J, Hughes CC (2004) Of mice and not men: differences between mouse and human immunology. J Immunol 172(5):2731–2738. https://doi.org/10.4049/jimmunol.172.5.2731
38. Puorro G, Marsili A, Sapone F, Pane C, De Rosa A, Peluso S, De Michele G, Filla A, Sacca F (2018) Peripheral markers of autophagy in polyglutamine diseases. Neurol Sci 39(1):149–152. https://doi.org/10.1007/s10072-017-3156-6
39. Areal LB, Pereira LP, Ribeiro FM, Olmo IG, Muniz MR, do Carmo Rodrigues M, Costa PF, Martins-Silva C, Ferguson SSG, Guimaraes DAM, Pires RGW (2017) Role of dynein axonemal heavy chain 6 gene expression as a possible biomarker for Huntington's disease: a translational study. J Mol Neurosci 63(3–4):342–348. https://doi.org/10.1007/s12031-017-0984-z
40. Castaldo I, De Rosa M, Romano A, Zuchegna C, Squitieri F, Mechelli R, Peluso S, Borrelli C, Del Mondo A, Salvatore E, Vescovi LA, Migliore S, De Michele G, Ristori G, Romano S, Avvedimento EV, Porcellini A (2018) DNA damage signatures in peripheral blood cells (PBMC) as biomarkers in prodromal Huntington's disease. Ann Neurol. https://doi.org/10.1002/ana.25393
41. Mastrokolias A, Ariyurek Y, Goeman JJ, van Duijn E, Roos RAC, van der Mast RC, van Ommen GB, den Dunnen JT, 't Hoen PAC, van Roon-Mom WMC (2015) Huntington's disease biomarker progression profile identified by transcriptome sequencing in peripheral blood. Eur J Hum Genet 23 (10):1349-1356https://doi.org/10.1038/ejhg.2014.281
42. Andrade-Navarro MA, Muhlenberg K, Spruth EJ, Mah N, Gonzalez-Lopez A, Andreani T, Russ J, Huska MR, Muro EM, Fontaine JF, Amstislavskiy V, Soldatov A, Nietfeld W, Wanker EE, Priller J (2020) RNA sequencing of human peripheral blood cells indicates upregulation of immune-related genes in Huntington's disease. Front Neurol 11:573560. https://doi.org/10.3389/fneur.2020.573560