Supplementary Material

A Multi-omic Huntington's Disease Transgenic Sheep-Model Database for Investigating Disease Pathogenesis

The following is a full description of each dataset included in the HD sheep database including experimental and data normalisation procedures, and previous publications.

Transcriptomic datasets from the 5-year-old cohort

RNA sequencing of striatum samples: Ref. data 1

The entire dorsal medial portion (averaging 207 mg) of the anterior striatum for each animal was sent to Expression Analysis (Morrisville, NC) for RNA extraction and RNA-Seq. Total RNA was isolated from multiple 30 mg portions encompassing the whole tissue portion using a RNeasy mini kit, as per the manufacturer's instructions (Qiagen). RNA quality was determined by Agilent 2100 Bioanalyzer (Agilent Technologies) and samples with an RNA Integrity (RIN) score >5 were pooled for each animal. A Ribo-Zero rRNA Removal Kit was used to deplete ribosomal RNAs. Twelve libraries were prepared using the TruSeq Stranded Total RNA kit and sequencing was conducted using the Illumina HiSeq2000 platform, generating a total of $2.3x10^9$ 50 bp paired end reads (all Illumina, San Diego, CA). Bioinformatic analysis was performed by AgResearch (Mosgiel, NZ), and included *de novo* transcriptome assembly with incorporation of Ensembl gene models for a newly available sheep genome (Texel breed) as it became available (Oar_v3.1; sheep release 75 feb2014.archive.ensembl.org/Ovis_aries/Info/Index). Data was provided as Fragments Per Kilobase of Exon Per Million Fragments Mapped (FPKM) counts for all transcripts, along with differential expression (DE) statistical results (comparing OVT73 to controls) in an Excel file. In this report, the dataset is referred to as Ref. data 1. Full details of the

study for Ref. data 1 and summary results were previously published by Handley and colleagues [1].

NanoString quantification of 24 genes (DM and DL portions of striatum): Ref. data 2

The nCounter gene expression analysis system (NanoString Technologies, WA) was used to re-quantify/validate 24 genes that were determined to be differentially expressed through the striatum RNA-Seq study. Custom CodeSet design and nCounter analysis was performed by New Zealand Genomics Ltd (NZGL; Dunedin, NZ). The dorsal medial portion of the striatum (DM) and the adjacent dorsal lateral portion of the striatum (DL) were analysed for each animal. RNA extractions were performed using the RNeasy mini kit (Qiagen) and quality assessments were made by Agilent 2100 Bioanalyzer (Agilent Technologies). Each sample was measured in duplicate, and results averaged. Transcript counts were normalised against four reference genes and internal positive control samples (nSolver software). Data was provided as normalised counts for the 24 named transcripts (duplicates averaged) in an Excel file. In this report, the dataset is referred to as Ref. data 2. Full details of the study for Ref. data 2 and summary results were previously published by Handley and colleagues [1].

RNA sequencing of laser-captured microdissected (LCM) matrix neurons in the striatum: Ref. data 3

Neurons from the matrix compartment of the 5-year-old-cohort striatal tissue were distinguished from striosomes through calbindin staining, which specifically targets the matrix. Matrix-derived neurons were isolated separately using LCM, resulting in a total of 12 matrix-derived neuronal samples. High quality RNA was extracted using the RNeasy Mini Kit (Qiagen)

and quantified via Qubit. Total RNA was sequenced on 3 lanes of an Illumina HiSeq 2000 by NZGL using TruSeq RNA libraries and collecting 100 bp paired end reads. Bioinformatics analysis was performed by AgResearch. All raw data was quality checked and filtered reads were mapped against the sheep reference genome (Ovis_aries_v3.1) using splice-aware mapper TopHap v2.0.12. Individual transcripts were quantified as FPKM and DE analysis, comparing transgenic matrix samples to control matrix samples, was done using Cufflinks v2.2.1. The LCM_Matrix Excel data file supplied, contains FPKM normalised counts for each named gene, for each of the 12 samples within the 5-year-old cohort, along with statistical results from DE analyses. In this report, the dataset is referred to as Ref. data 3. This data has not been previously published.

Supplementary Table 1. OVT73 and Control gene expression data measured by three methods (striatal RNAseq, selected gene nanostring RNA quantification and the LCM Striatal neurone RNA-Seq) nominal significance values and direction of change OVT73 v's controls.

	RNA-Seq		nanoString RNA quantification		LCM Striatal Neuron RNA-Seq	
Gene	Significance (nominal p)	Direction of change in Transgenics	Significance (nominal p)2	Direction of change in Transgenics2	Significance (nominal p)4	Direction of change in Transgenics5
AP2S1	0.454	NC	0.0902 (Dorsolateral) 0.321 (Dorsomedial)	NC NC	0.424	NC
AVPR1A	0.119	NC	0.214 (Dorsolateral) 0.121 (Dorsomedial)	NC NC	0.395	NC
CBS	0.022	Up	0.24 (Dorsolateral) 0.0155 (Dorsomedial)	NC Up	0.725	NC
CHST8	0.338	NC	0.138 (Dorsolateral) 0.267 (Dorsomedial)	NC NC	0.369	NC
CMTM5	0.383	Down	0.561 (Dorsolateral) 0.139 (Dorsomedial)	NC NC	0.079	NC
CNTN2	0.569	NC	0.245 (Dorsolateral) 0.408 (Dorsomedial)	NC NC	0.830	NC
CPAMD8	0.015	Down	0.775 (Dorsolateral) 0.0477 (Dorsomedial)	NC Down	0.971	NC
ENS11790	0.000	Up	0.265 (Dorsolateral) 0.531 (Dorsomedial)	NC NC	Not Detected	
ETV5	0.047	Up	0.592 (Dorsolateral) 0.0207 (Dorsomedial)	NC Up	0.243	NC
FEZF2	0.129	NC	0.887 (Dorsolateral) 0.207 (Dorsomedial)	NC NC	0.158	NC
HPN	0.038	Down	0.841 (Dorsolateral) 0.0421 (Dorsomedial)	NC Down	0.313	NC
HSD17B12-like	0.632	NC	0.0842 (Dorsolateral)	NC	0.793	NC

			0.06 (Dorsomedial)	NC		
ITGB4	0.004	Down	0.404 (Dorsolateral)	NC	0.105	NC
			0.0152 (Dorsomedial)	Down		
MYL4	0.484	NC	0.265 (Dorsolateral)	NC	0.952	NC
			0.476 (Dorsomedial)	NC		
OTX2	0.034	Up	0.984 (Dorsolateral)	NC	0.528	NC
			0.614 (Dorsomedial)	NC		
OXTR	0.003	Up	0.220 (Dorsolateral)	NC	0.439	NC
			0.0131 (Dorsomedial)	Up		
PRKG1	0.057	NC	0.127 (Dorsolateral)	NC	0.164	NC
			0.833 (Dorsomedial)	NC		
RHCG	0.105	NC	0.405 (Dorsolateral)	NC	0.443	NC
			0.047 (Dorsomedial)	Up		
SIAH3	0.008	Up	0.212 (Dorsolateral)	NC	0.549	NC
			0.00115 (Dorsomedial)	Up		
SLC14A1	0.001	Up	0.00540 (Dorsolateral)	Up	0.003	Up
			0.00698 (Dorsomedial)	Up		
SLC5A7	0.025	Up	0.276 (Dorsolateral)	NC	0.272	NC
			0.00825 (Dorsomedial)	Up		
SMOC2	0.011	Up	0.149 (Dorsolateral)	NC	0.657	NC
			0.0274 (Dorsomedial)	Up		
TF	0.398	NC	0.949 (Dorsolateral)	NC	0.230	NC
			0.106 (Dorsomedial)	NC		
ZNF804A	0.133	NC	0.231 (Dorsolateral)	NC	0.749	NC
			0.15 (Dorsomedial)	NC		
ovineHTT	0.748	NC	0.95 (Dorsolateral)	NC	0.354	NC
			0.039 (Dorsomedial)	Up		

This table represents the 25 genes as measured by three methods of RNA quantification (RNASeq from bulk tissue, nanoString, and RNASeq from LCM neurons), comparing the change observed in OVT73 sheep compared to their control counterparts. The p-value presented is nominal from a two-tailed t-test without false-discovery rate correction. A direction of change was entered if nominal significance was < 0.05. Note: ENS11790 was not measured in the LCM data set.

Supplementary Table 2. Nominally significant genes from the Bulk tissue RNASeq and LCM RNASeq data sets.

Gene ID	Change in OVT73 RNASeq Bulk tissue assay	RNASeq Bulk tissue assay P-value	Change in OVT73 LCM Striatal Neuron RNASeq	LCM Striatal Neuron RNASeq p
ALDH9A1	Decrease	0.0122	Decrease	0.0494
AMP18	Increase	0.0296	Increase	0.0382
ATOH8	Decrease	0.0092	Increase	0.0319
EPB41L4A	Decrease	0.0295	Decrease	0.0014
FTCDNL1	Decrease	0.0176	Decrease	0.0441
GRK5	Increase	0.0261	Increase	0.0281
HKDC1	Decrease	0.0354	Decrease	0.0268
LMO7	Increase	0.0222	Increase	0.0317
LRRC52	Increase	0.0410	Increase	0.0303
RBP1	Increase	0.0176	Increase	0.0298
SERPINB9	Increase	0.0083	Increase	0.0116
SI	Increase	0.0374	Decrease	0.0376
SLC14A1	Increase	0.0015	Increase	0.0088
TM2D3	Increase	0.0122	Increase	0.0377
UCK2	Increase	0.0451	Increase	0.0035
YBEY	Decrease	0.0150	Decrease	0.0492

All nominally significant genes from the Bulk striatal tissue RNASeq and LCM RNASeq data sets are listed in the table. This is presented as a comparison of the two RNASeq experiments to identify the genes that are nominally significant different in expression in OVT73 sheep compared to their control counterparts. (Two-sided t-test p-value < 0.05).

Metabolic data from the 5-year-old cohort

Local metabolite assessment of cerebellum, hippocampus, motor cortex and liver samples: Ref. data 4

Local gas chromatography-mass spectrometry (GC-MS) analysis of ~50 metabolites was performed by The University of Auckland Centre for Genomics, Proteomics and Metabolomics (CGPM). Approximately 300 mg of fresh-frozen cerebellum, hippocampus, motor cortex, and liver tissue was crushed into a fine powder under liquid nitrogen. Metabolic processes were quenched, and small polar metabolites extracted using a cold methanol-based method. A deuterium labelled internal standard 2,3,3,3-d4-alanine (0.4 μM) was added to each sample

immediately before metabolite extraction. Samples were split into two aliquots and a methyl-chloroformate method was used to extract metabolites from each technical replicate. The methyl-chloroformate derivatives were assessed using an Agilent 7890A gas chromatograph coupled to a 5975C inert mass spectrometer. Following GC-MS, metabolites were identified and quantified using Metab, an R package for automated analysis of GC-MS data. The raw abundance of each named metabolite was determined as the height of its designated major mass fragment, before being normalised to the abundance of the internal standard and to the weight of the dried tissue pellet for the same sample. Data was provided as relative abundance values for each metabolite, log10 transformed to approximate a normal distribution, and provided as an excel file. In this report, the dataset is referred to as Ref. data 4. A detailed description of the study for Ref. data 4 and summary results were previously published by Handley and colleagues [2].

Biocrates metabolite assessment of motor cortex, cerebellum, liver, and plasma samples: Ref. data 5

Quantification of 180 metabolites within motor cortex, cerebellum, liver, and plasma samples was undertaken by Biocrates Life Sciences AG (Innsbruck, Austria), using the AbsoluteIDQ p180 Kit assay. Approximately 50-100 mg of each solid tissue sample (and 600 μL of plasma) was sent to Biocrates for the quantification of amino acids, acylcarnitines, sphingomyelins, phosphatidylcholines, hexoses, and biogenic amines. The assay uses PITC (phenylisothiocyanate)-derivatization in the presence of internal standards followed by flow injection analysis–tandem mass spectrometry (FIA-MS-MS) (acylcarnitines, lipids, and hexose) and liquid chromatography-mass spectrometry (LC-MS) (amino acids, biogenic amines) using an AB SCIEX 4000 QTrap™ mass spectrometer (AB SCIEX, Darmstadt, Germany) with

electrospray ionization. The experimental metabolomics measurement technique is described in detail by patent US 2007/0004044 (accessible online at http://www.freepatentsonline.com/y2007/0004044.html). Biocrates "LOD" files were generated, where only abundances above the lower limit of detection were reported. Data for tissues is expressed as pmol/mg solid tissue and μM for plasma. Datasets were log10 transformed to approximate a normal distribution and provided as an Excel file. In this report, the dataset is referred to as Ref. data 5. This data has not been previously published.

Proteomic data from the 5-year-old cohort: Ref. data 6

Protein abundance quantification was achieved using liquid chromatography-tandem mass spectrophotometry (LC-MS-MS) coupled with isobaric mass tagging (iTRAQ) at the Centre for Advanced Discovery and Experimental Therapeutics (CADET), University of Manchester, UK. Striatum, cerebellum, and motor cortex samples were collected from fresh frozen tissue at the University of Auckland. Proteins were extracted, and quality control (QC) checked before being shipped on dry ice to CADET. LC-MS-MS coupled with iTRAQ was conducted on a total of 36 brain samples (12 per tissue). Experimental methods were as described in [3]. Data were searched against an NCBI Ovis-specific database (version: Oar_V3.1; 20833 protein coding; download date: July 2013) using ProteinPilot software (v4, Sciex). Relative peptide and protein abundance for each sample was quantified by determining the height of the peaks from iTRAQ reporter mass intensities. The abundance of all peptides matching a specific protein were combined to estimate the relative abundance of each protein in each sample. Appropriate false discovery rate scores were applied to ensure reliable identification of the proteins, as described in [4]. Data was provided as the relative protein abundance for each named protein in an Excel file,

with separate Excel sheets for each of the three tissue regions quantified. In this report, the dataset is referred to as Ref. data 6. This data has not been previously published.

Follow-up data from the 5-year-old cohort

Quantification of urea within tissue and fluid samples: Ref. data 7

Biochemical quantification of urea was performed on a range of tissue samples obtained from the 5-year-old cohort (cerebellum, hippocampus, motor cortex, striatum, bladder, heart, kidney, liver, testes, serum, and urine) using the Urea Assay Kit (ab83362) as per the manufacturer's instructions (Abcam, Cambridge, UK). Sample extracts were assayed in triplicate and background controls assayed in duplicate. The protein content in each sample lysate was determined by Bio-Rad DC protein assay (Bio-Rad, Auckland, NZ). Urea data from 12 animals (across 11 tissues) was presented as nmol urea per mg protein and provided as an Excel file. In this report, the dataset is referred to as Ref. data 7. Details of the study for Ref. Data 7 and summary results were previously published by Handley and colleagues [1].

REFERENCES

[1] Handley RR, Reid SJ, Brauning R, Maclean P, Mears ER, Fourie I, et al. Brain urea increase is an early Huntington's disease pathogenic event observed in a prodromal transgenic sheep model and HD cases. Proc Natl Acad Sci U S A. 2017;114(52):E11293-E302. doi:10.1073/pnas.1711243115.

[2] Handley RR, Reid SJ, Patassini S, Rudiger SR, Obolonkin V, McLaughlan CJ, et al. Metabolic disruption identified in the Huntington’s disease transgenic sheep model. Sci Rep. 2016;6:20681. doi:10.1038/srep20681.

[3] Xu J, Patassini S, Rustogi N, Riba-Garcia I, Hale BD, Phillips AM, et al. Regional protein expression in human Alzheimer's brain correlates with disease severity. Commun Biol. 2019;2(1):43. doi:10.1038/s42003-018-0254-9.

[4] Patassini S. Discovery and validation of relevant markers of Huntington's disease progression using a transgenic sheep model. Auckland, New Zealand: The University of Auckland (PhD Thesis); 2014.