Supplementary Information for "When do mouse models transfer? Cross-species prediction of Huntington's disease from brain transcriptomes"

S1. Supplementary Methods

Exploratory family and multiplicity. The striatum→caudate analysis (all mice, Q20 controls, LR) is the only confirmatory test. All other transfer tests form one exploratory family of 32 tests: five region-grid cells and the BA9 cohort, composition-invariant transfer, six sorted-nucleus targets and the MSN-minus-interneuron contrast, and 18 benchmark tests (nine contrasts × two human targets). Benjamini–Hochberg q-values were computed over their one-sided mouse-label permutation p-values (1,000 permutations each). The staging map (60 cells) was corrected within itself.

Sorted nuclei. GSE227729 contains HD samples only for neurons (D1 and D2 MSNs and interneurons). Controls were also sampled from posterior caudate and putamen, so "matched" analyses used only anterior caudate, anterior putamen, and caudate. Scores were averaged per donor before computing AUC; permutation and random-gene nulls were scored identically. HD donors had shorter postmortem intervals; scores did not track postmortem interval among controls ($\rho = -0.13$ to -0.25).

Simulation. Target: 70 samples (38 HD, 32 controls), 1,000 genes; a latent factor loads 30% of genes with loading 0.7 and is shifted by $\Delta \in \{0, 1, 2, 3\}$ in HD. Source: 64 samples with labels independent of all features; source loadings equal the target's with probability 0 or 0.8. The pipeline used $k = 100$ genes; 150 replicates per setting; false-positive rate at $\alpha = 0.05$.

PCA-projection baseline. Ten principal components were learned without labels from either the mouse training data or the human target data; both species were projected onto them, LR ($C = 0.1$, balanced) was trained on the mouse projections and applied unchanged to the human projections, and significance was assessed with 1,000 within-age mouse-label permutations.

Software. Fifteen Python analysis scripts (scikit-learn, SciPy, pandas) with fixed random seeds generated every number, table and figure reported. Running them requires downloading the public GEO datasets listed in Table 1 of the main text. Figures for the batch audit, the staging map and the calibration of the analytic null appear in the main text (Figures 8, 9 and 5).

Table S1: Human AUC by training design (LR with 95% CI; RF point estimate). Mouse CV: within-mouse cross-validated AUC (LR).

Analysis	Design	Controls	Mouse CV	Human AUC, LR	RF
Striatum→caudate	all	Q20	1.00	0.87 (0.77–0.95)	0.85
Striatum→caudate	all	WT+Q20	1.00	0.83 (0.72–0.92)	0.75
Striatum→caudate	within-batch	WT+Q20	1.00	0.93 (0.86–0.98)	0.76
Striatum→caudate	within-batch	Q20	1.00	0.90 (0.81–0.96)	0.73
Cortex→BA9	all	Q20	0.95	0.63 (0.48–0.76)	0.69
Cortex→BA9	all	WT+Q20	0.99	0.59 (0.45–0.73)	0.58
Cortex→BA9	within-batch	WT+Q20	0.82	0.49 (0.34–0.63)	0.55
Cortex→BA9	within-batch	Q20	0.66	0.40 (0.25–0.55)	0.35

Table S2: Region grid and BA9 (LR, all mice, Q20 controls). p_{MW}: one-sided Mann–Whitney over human samples; p_{perm}, p_{rand}: 1,000 mouse-label permutations and random 200-gene sets; q: BH over the exploratory family.

Mouse	Human	AUC	p_{MW}	p_{perm}	p_{rand}	q
Striatum	caudate	0.87	7×10^{-8}	0.035	0.12	confirmatory
Cortex	caudate	0.88	4×10^{-8}	0.031	0.031	0.15
Striatum	frontal	0.50	0.51	0.52	0.31	0.67
Cortex	frontal	0.59	0.11	0.21	0.20	0.40
Striatum	cerebellum	0.34	0.98	0.91	0.79	0.92
Cortex	cerebellum	0.66	0.015	0.095	0.039	0.26
Cortex	BA9 (GSE64810)	0.63	0.051	0.29	0.23	0.43

Table S3: Transfer inside sorted human striatal nuclei (donor level). HD/C: donors.

Target	HD/C	AUC (LR)	AUC (RF)	p_{perm}	p_{rand}	τ	q
MSNs, all regions	5/8	0.95	1.00	0.051	0.006	1.06	0.20
MSNs, matched	5/8	0.80	1.00	0.13	0.021	1.20	0.28
D1 MSNs, matched	4/8	0.81	0.97	0.10	0.040	1.26	0.26
D2 MSNs, matched	5/7	0.83	1.00	0.11	0.009	1.10	0.26
Interneurons, all	4/8	0.28	0.88	0.76	0.092	0.76	0.81
Interneurons, matched	4/3	0.25	0.83	0.74	0.068	1.01	0.81
MSN − interneuron		+0.67		0.027	0.14		0.15

Table S4: Mouse benchmark: human AUC (LR) with permutation p and BH q. DE: mouse genes with $|t| > 3$ (caudate gene set). Cohort 1: Pms1 study; cohort 2: Msh3 study.

Contrast	n	DE	Caudate	Sorted MSNs
zQ175 vs. WT	8/8	2,786	0.94 (0.012; q 0.13)	0.97 (0.018; q 0.14)
R6/2 vs. WT (TRAP)	11/11	2,947	0.97 (1×10^{-3}; q 0.032)	0.80 (0.11; q 0.26)
Q140 vs. WT, cohort 1	8/8	3,285	0.89 (0.061; q 0.22)	0.90 (0.086; q 0.26)
Q140 vs. WT, cohort 2	8/7	2,465	0.36 (0.63; q 0.74)	0.75 (0.23; q 0.41)
Q140 vs. Q140;Pms1$^{-/-}$	8/8	1,174	0.95 (0.003; q 0.048)	0.40 (0.62; q 0.74)
Q140 vs. Q140;Msh3$^{-/-}$	8/8	1,780	0.47 (0.50; q 0.67)	0.70 (0.30; q 0.43)
zQ175 vs. zQ175;Msh3$^{-/-}$ (ctrl)	8/8	59	0.84 (0.20; q 0.40)	0.70 (0.27; q 0.43)
WT vs. Msh3$^{-/-}$ (ctrl)	7/8	65	0.75 (0.25; q 0.42)	0.62 (0.41; q 0.57)
WT vs. Pms1$^{-/-}$ (ctrl)	8/8	878	0.24 (0.74; q 0.81)	0.15 (0.92; q 0.92)

Table S5: Rescue check: primary allelic-series model scores in independent mouse cohorts (mean ± SD).

Cohort	Group	n	Score
Q140×Msh3	Q140	8	0.99 ± 0.02
Q140×Msh3	Q140;Msh3KO	8	0.23 ± 0.29
Q140×Msh3	Q140;Msh3het	8	0.74 ± 0.33
Q140×Msh3	Msh3KO	8	0.00 ± 0.00
Q140×Msh3	Msh3het	9	0.00 ± 0.00
Q140×Msh3	WT	7	0.01 ± 0.00
Q140×Pms1	Pms1het	8	0.00 ± 0.00
Q140×Pms1	WT	8	0.00 ± 0.00
Q140×Pms1	Q140;Pms1het	8	0.98 ± 0.01
Q140×Pms1	Q140	8	1.00 ± 0.00
Q140×Pms1	Pms1KO	8	0.00 ± 0.00
Q140×Pms1	Q140;Pms1KO	8	0.54 ± 0.26
zQ175×Msh3	Msh3KO	8	0.00 ± 0.00
zQ175×Msh3	WT	8	0.00 ± 0.00
zQ175×Msh3	zQ175;Msh3KO	8	0.98 ± 0.01
zQ175×Msh3	zQ175	8	0.98 ± 0.00
R6/2 TRAP	CTRL	11	0.00 ± 0.00
R6/2 TRAP	R62	11	0.91 ± 0.05

Table S6: PCA-projection baseline (TransComp-R-style) versus the primary pipeline, striatum→caudate.

Model	AUC	p_{perm}	Null 97.5%
PCA projection (PCs from mouse, 10 PCs)	0.83	0.30	0.94
PCA projection (PCs from human, 10 PCs)	0.61	0.32	0.82
Primary pipeline (ANOVA-200 + LR)	0.87	0.035	0.88

S2. Robustness analyses

Reference-based deconvolution. Reference profiles for five classes were averaged from sorted control nuclei (GSE227729); the 50 most specific genes per class with mean $\geq$ 10 CPM (250 markers) formed the signature. Microarray values were linearized, each marker scaled by its reference mean, samples median-scaled, and proportions estimated by non-negative least squares and renormalized. The reference lacks endothelial cells and oligodendrocyte precursors and comes from nuclear RNA on a different platform. The projection-neuron fraction alone separated HD from controls with AUC 0.96 and correlated with the transfer score ($\rho = -0.63$).

Table S7: Deconvolved cell-type proportions in human caudate (GSE3790), mean per group; Mann–Whitney p.

Cell class	Control	HD	p
MSN	0.46	0.22	6×10^{-11}
Interneuron	0.11	0.12	0.10
Astrocyte	0.26	0.42	7×10^{-9}
Microglia	0.05	0.07	1×10^{-4}
Oligodendrocyte	0.12	0.17	9×10^{-5}

Table S8: Sensitivity of the primary analysis (LR; 1,000 mouse-label permutations). Permutation p-values vary across independent runs by Monte Carlo error (primary: 0.018–0.035).

Variant	Genes	AUC	p_{perm}	Null 97.5%
primary (HCOP support $\geq$ 3)	7,780	0.87	0.018	0.86
HCOP support $\geq$ 11	6,986	0.84	0.061	0.88
HCOP support = 12	4,743	0.88	0.015	0.87
within-sample rank-INT	7,780	0.95	0.001	0.85

Table S9: Baselines and adjustments for the primary analysis, each with its own mouse-label permutation null.

Model	AUC	p_{perm}	Null 97.5%
PCA projection (PCs from mouse, 10 PCs)	0.83	0.30	0.94
PCA projection (PCs from human, 10 PCs)	0.61	0.32	0.82
CORAL + LR	0.76	0.081	0.83
Random forest (200 trees; 200 permutations)	0.85	0.085	0.89
Score adjusted for deconvolved proportions	0.47	0.77	0.61

Table S10: Between-cohort heterogeneity of HD-vs-WT transfer (four mouse cohorts) versus the mean Hanley–McNeil sampling error of each AUC.

Target	Cohorts	Mean AUC	Between-cohort SD	Sampling SE	Ratio
Caudate	4	0.79	0.29	0.04	7.4
Sorted MSNs	4	0.86	0.10	0.11	0.9

S3. Calibration of the analytic null

Each replicate simulated a target (70 samples, 400 genes, latent factor shifted by Δ in HD) and a 64-sample source. Without signal, source labels were independent of the source data; with signal, they followed a source factor whose loadings matched the target disease factor. Each replicate recorded the naive Mann–Whitney p, the analytic p, and a permutation p (99 permutations); 100 replicates per setting. Grouping all no-signal replicates by τ: below 0.4 the naive false-positive rate was 0.03; at 0.4–0.8 and above 0.8 it was 0.27 and 0.29, while analytic and permutation rates stayed at 0.01–0.03.

Table S11: Rejection rates at $\alpha = 0.05$ (false-positive rate without signal; power with signal).

Scenario	Δ	Sharing	Median τ	Naive	Analytic	Permutation
No signal	0.0	0.0	0.23	0.01	0.03	0.01
No signal	1.5	0.0	0.56	0.27	0.02	0.03
No signal	3.0	0.0	0.87	0.33	0.02	0.02
No signal	0.0	0.8	0.20	0.04	0.04	0.03
No signal	1.5	0.8	0.98	0.22	0.00	0.04
No signal	3.0	0.8	1.69	0.34	0.00	0.03
Signal	1.5	0.8	0.99	1.00	0.14	1.00
Signal	3.0	0.8	1.71	1.00	0.77	1.00