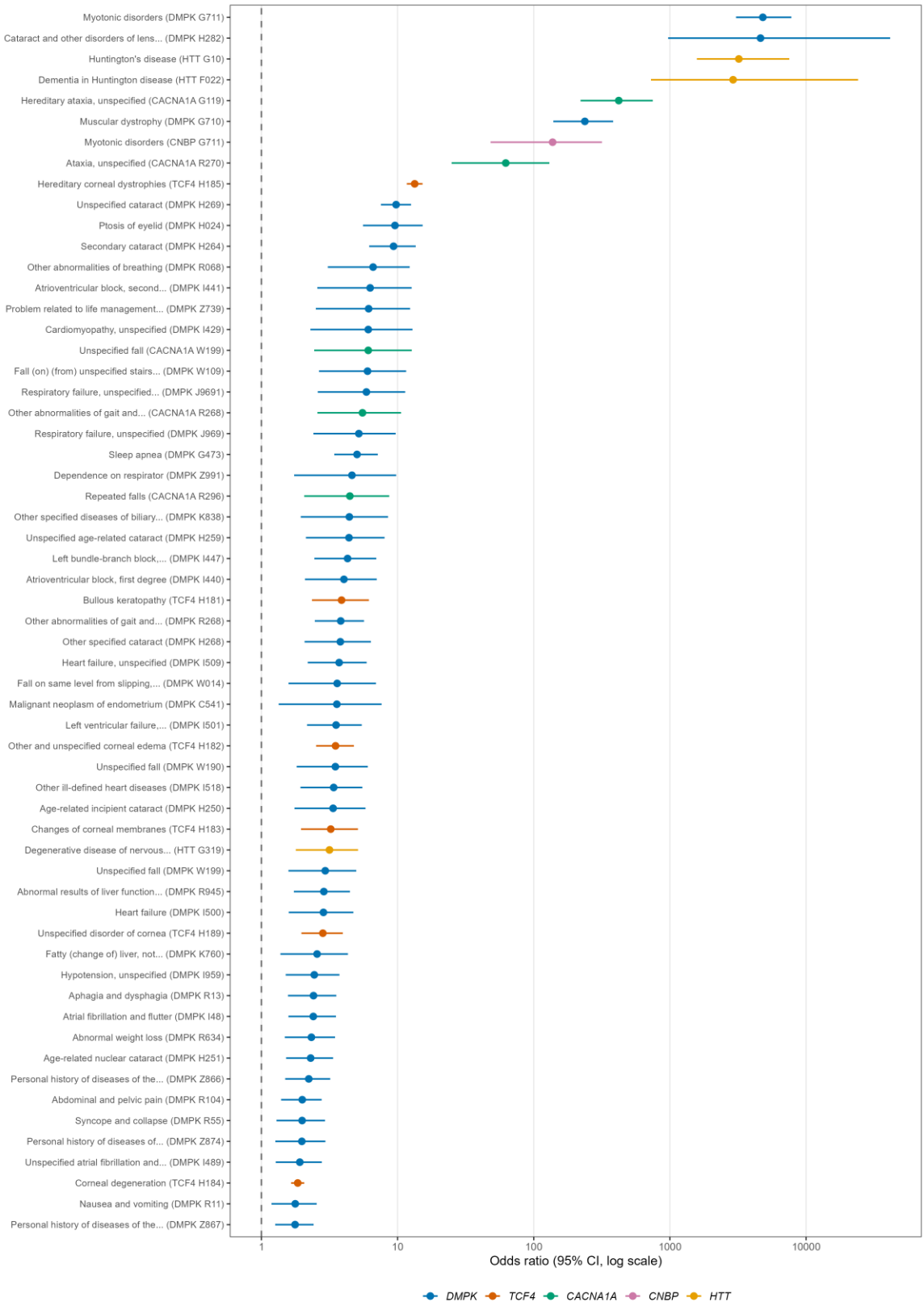
Supplementary Information

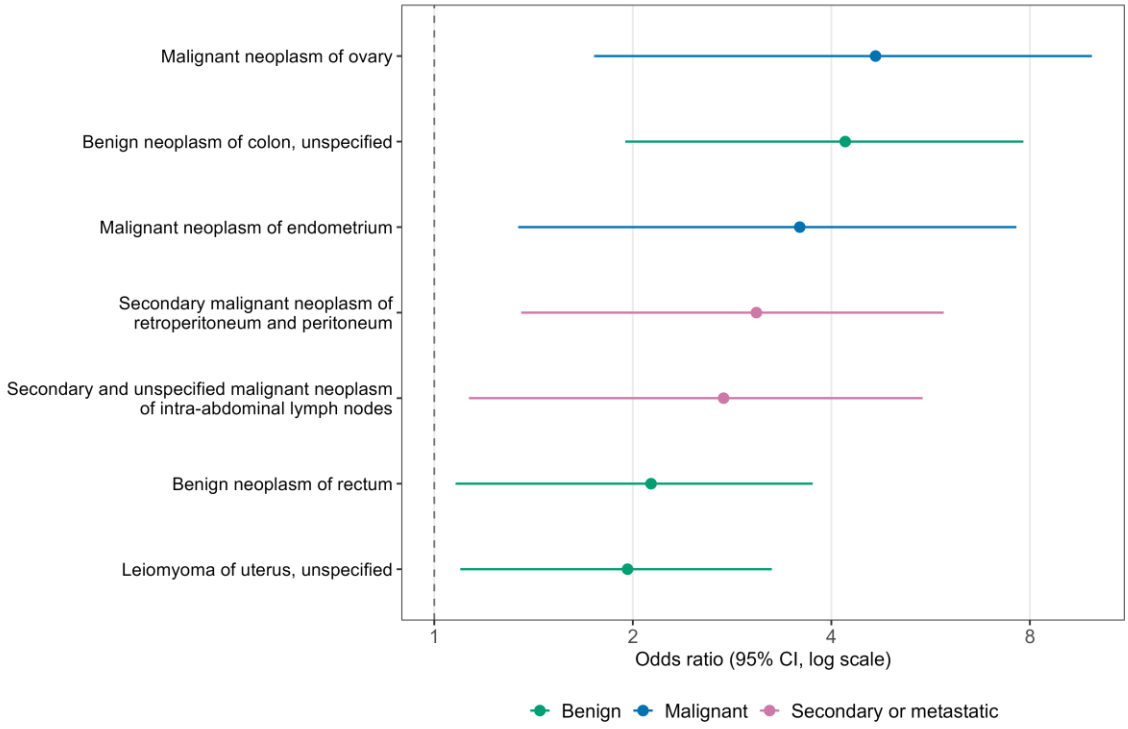
Phenome wide analysis of neurological disease associated short tandem repeats in UK Biobank

Supplementary Figure S1


Myotonic disorders (DMPK G711)
Cataract and other disorders of lens... (DMPK H282)
Huntington's disease (HTT G10)
Dementia in Huntington disease (HTT F022)
Hereditary ataxia, unspecified (CACNA1A G119)
Muscular dystrophy (DMPK G710)
Myotonic disorders (CNBP G711)
Ataxia, unspecified (CACNA1A R270)
Hereditary corneal dystrophies (TCF4 H185)
Unspecified cataract (DMPK H269)
Ptosis of eyelid (DMPK H024)
Secondary cataract (DMPK H264)
Other abnormalities of breathing (DMPK R068)
Atrioventricular block, second... (DMPK I441)
Problem related to life management... (DMPK Z739)
Cardiomyopathy, unspecified (DMPK I429)
Unspecified fall (CACNA1A W199)
Fall (on) (from) unspecified stairs... (DMPK W109)
Respiratory failure, unspecified... (DMPK J9691)
Other abnormalities of gait and... (CACNA1A R268)
Respiratory failure, unspecified (DMPK J969)
Sleep apnea (DMPK G473)
Dependence on respirator (DMPK Z991)
Repeated falls (CACNA1A R296)
Other specified diseases of biliary... (DMPK K838)
Unspecified age-related cataract (DMPK H259)
Left bundle-branch block,... (DMPK I447)
Atrioventricular block, first degree (DMPK I440)
Bullous keratopathy (TCF4 H181)
Other abnormalities of gait and... (DMPK R268)
Other specified cataract (DMPK H268)
Heart failure, unspecified (DMPK I509)
Fall on same level from slipping,... (DMPK W014)
Malignant neoplasm of endometrium (DMPK C541)
Left ventricular failure,... (DMPK I501)
Other and unspecified corneal edema (TCF4 H182)
Unspecified fall (DMPK W190)
Other ill-defined heart diseases (DMPK I518)
Age-related incipient cataract (DMPK H250)
Changes of corneal membranes (TCF4 H183)
Degenerative disease of nervous... (HTT G319)
Unspecified fall (DMPK W199)
Abnormal results of liver function... (DMPK R945)
Heart failure (DMPK I500)
Unspecified disorder of cornea (TCF4 H189)
Fatty (change of) liver, not... (DMPK K760)
Hypotension, unspecified (DMPK I959)
Aphagia and dysphagia (DMPK R13)
Atrial fibrillation and flutter (DMPK I48)
Abnormal weight loss (DMPK R634)
Age-related nuclear cataract (DMPK H251)
Personal history of diseases of the... (DMPK Z866)
Abdominal and pelvic pain (DMPK R104)
Syncope and collapse (DMPK R55)
Personal history of diseases of... (DMPK Z874)
Unspecified atrial fibrillation and... (DMPK I489)
Corneal degeneration (TCF4 H184)
Nausea and vomiting (DMPK R11)
Personal history of diseases of the... (DMPK Z867)
1
10
100
1000
10000
Odds ratio (95% CI, log scale)
DMPK
TCF4
CACNA1A
CNBP
HTT

Supplementary Figure S1. Forest plot of 59 established associations, coloured by locus. ORs and 95% CIs use logarithmic scales and age-, sex- and GPC1-10-adjusted models.

Supplementary Figure S2



Supplementary Figure S2. Seven nominally significant *DMPK* neoplasm associations among 15 eligible codes. ORs and 95% CIs use age-, sex- and GPC1-10-adjusted models. Colours distinguish malignant, benign and secondary/metastatic neoplasms.

Supplementary Tables

Supplementary Tables S1 to S9 are provided as a separate Excel workbook (Supplementary_Tables_S1-S9.xlsx), one table per sheet. Legends follow.

Supplementary Table S1. Genotyped cohort characteristics (n = 489,727): age, genetic and self-reported sex, ethnicity, GPC completeness, hospital diagnoses, recruitment, centres, follow-up and deaths. The primary analysis excludes 5,166 participants without GPCs or genetic sex. The wider phenotype release contains 12,504 additional participants without genotypes; whole-release figures are identified separately.

Supplementary Table S2. Catalogue of 49 STR entries: 48 analysed variants at 41 loci and excluded *C9orf72*. Seven alternative-motif or adjacent-tract entries are flagged. Columns report motif, associated disease, pathogenic threshold and carrier count.

Supplementary Table S3. ExpansionHunter Classifier results by locus, including pathogenic-range call counts and the proportion classified as true expansions. Uncovered loci, including *TCF4* and *CBL*, are absent.

Supplementary Table S4. Curated cancer groupings used for the pooled neoplasm endpoints: each ICD-10 code with its major category and subcategory, whether it defined a case, the scope of any control exclusion and whether it counted only towards the all-cancer endpoint, with endpoint definitions including all malignant neoplasms (C00 to C97 excluding C44). Hospital inpatient ICD-10 diagnoses only. Category names follow the source scheme.

Supplementary Table S5. FDR-significant associations not classified as established, separated into downstream, unrecognised and unresolved relationships. Models adjust for age, sex and GPCs 1 to 10 in 484,561 participants and require five carrier cases.

Supplementary Table S6. Established associations from age-, sex- and GPC1-10-adjusted models in 484,561 participants.

Supplementary Table S7. All 13,191 eligible tests at 19 variants, adjusted for age, sex and GPCs 1 to 10 in 484,561 participants, including ORs, CIs, P values, within-variant FDR and significance status. Eligibility requires at least one case, one non-case and five carrier cases. Cohort counts refer to the analysis cohort; the exact carrier-case column refers to all 489,727 genotyped participants. *C9orf72* is excluded; *ATXN7* has no eligible diagnosis. CI

methods are specified; zero P values indicate numerical underflow. ICD-10-CM descriptions are used where available; 483 WHO-only codes lack equivalents.

Supplementary Table S8. Allele-coding sensitivity analyses for 144 primary autosomal associations. (a) Pooled, single- and double-expanded groups versus non-carriers, and two versus one expanded allele within carriers. Models adjust for age, sex and GPCs 1 to 10; estimates include ORs, profile-likelihood 95% CIs, P values and within-locus FDR. (b) ExpansionHunter REPCN allele-group counts. Two-allele estimates require five double-expanded cases. X-linked *AR*, *ARX* and *FMR1* are excluded.

Supplementary Table S9. Within-carrier repeat-length analyses at 11 autosomal loci, adjusted for age, sex and GPCs 1 to 10, requiring five carrier cases and applying within-locus FDR correction. Counts refer to carriers with complete GPCs. ORs are reported per repeat and per ten repeats; extrapolation and inferred lengths are flagged.