Frequency and phenotypic characterization of genetic causes of sporadic late-onset cerebellar ataxia: insights from a cohort of 307 patients

Supplementary

List of the 392 genes included in our Parkinson, Movement Disorders and Ataxia-related genes panel

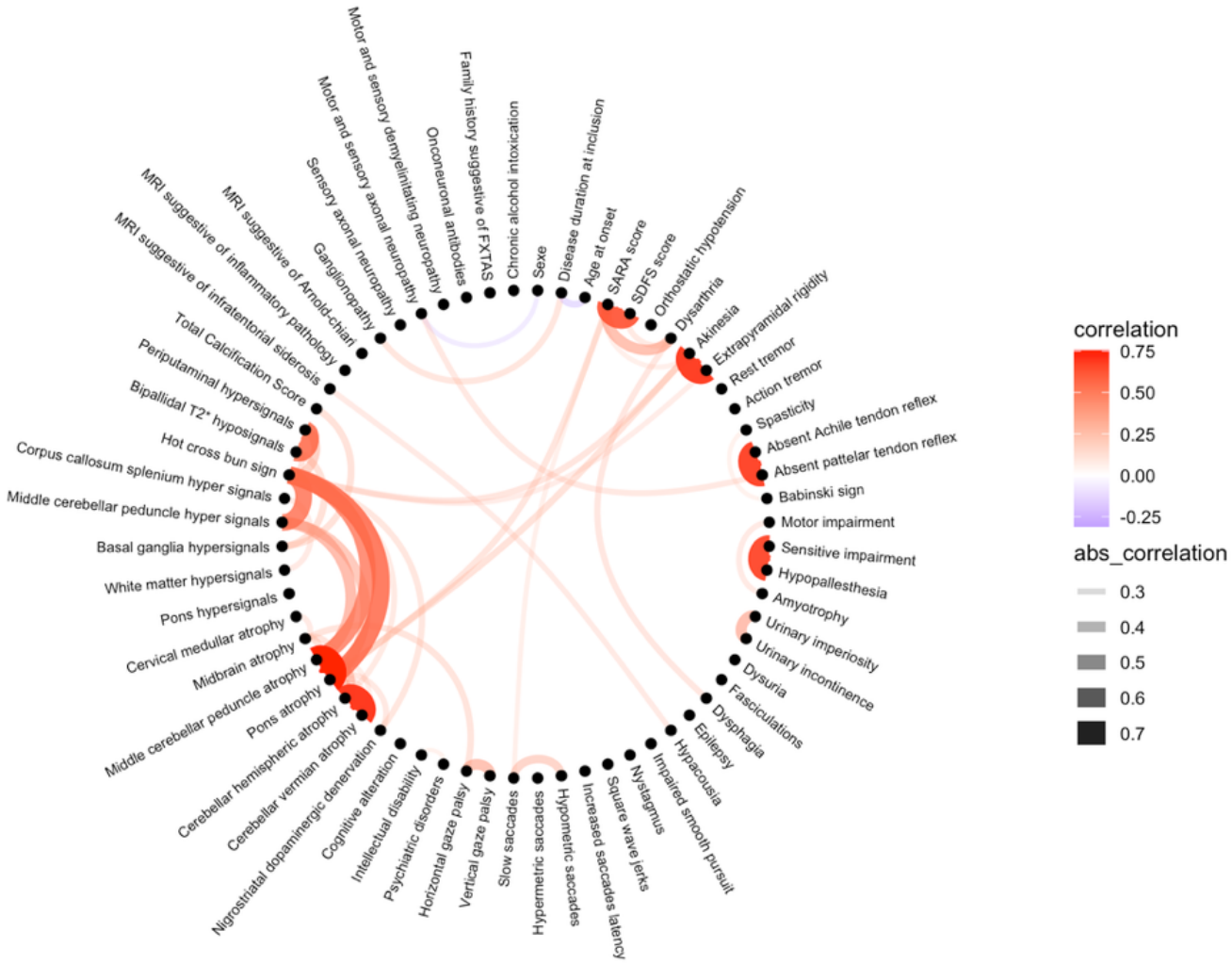
AAAS; ABCB7; ABCD1; ABHD12; ACO2; ACTB; ADCK3; ADCY5; AF3GL2; AHI1; ALAS2; ALDH18A1; ALDH5A1; ALDH7A1; ALG13; ALG6; AMACR; ANKRD11; ANO10; ANO3; AP4B1; AP5Z1; APTX; ARL13B; ARSA; ARX; ATAD3A; ATCAY; ATL1; ATM; ATP13A2; ATP1A3; ATP6AP2; ATP7B; ATP8A2; ATXN10; AUH; B4GALNT1; BCKDHA; BCKDHB; BEAN1; BTD; C10orf2; CA8; CACNA1A; CACNA1B; CACNA1E; CACNA1G; CACNB4; CAPN1; CBS; CC2D2A; CCDC88C; CDKL5; CEP290; CHAC1; CHCHD10; CHCHD2; CHP1; CIZ1; CLCN2; CLN3; CLN5; CLN6; CLTC; CNTNAP1; COA7; COASY; COL4A1; COL4A1; COL6A3; COMT; COQ2; COX15; COX20; CP; CSF1R; CSTB; CTNNB1; CTSF; CWF19L1; CYP27A1; CYP7B1; DARS2; DBT; DCAF17; DCC; DCTN1; DDC; DLAT; DLT; DNAJC12; DNAJC13; DNAJC19; DNAJC3; DNAJC5; DNAJC6; DNM1L; DNMT1; DNMT3A; DRD5; ECHS1; EEF2; EIF2B1; EIF2B2; EIF2B3; EIF2B4; EIF2B5; EIF4G1; ELOVL4; ELOVL5; EPM2A; ERCC2; ERCC3; ERCC4; ERCC5; ETHE1; FA2H; FAT3; FBXO11; FBXO7; FGF14; FLVCR1; FOLR1; FOXG1; FOXRED1; FRRS1L; FTL; FUCA1; FXN; GALC; GAMT; GAN; GATM; GBA; GBA2; GBE1; ; GCDH; GCH1; GDAP2; GFAP; GIGYF2; GJB1; GJC2; GLB1; GLRA1; GLRB; GM2A; GNAL; GNAO1; GOSR2; GPR56; GRIA3; GRID2; GRIN1; GRIN2A; GRIN2B; GRM1; GRN; HEXA; HEXB; HIBCH; HPCA; ; HPRT1; HTRA2; HUWE1; INPP5E; IREB2; ITPR1; KCNA1; KCNA2; KCNB1; KCNC1; KCND3; KCNJ10; KCNJ6; KCNMA1; KCNQ2; KCTD17; KCTD7; KIAA02226; KIF1C; KIF5A; KMT2B; L2HGDH; LAMA1; LMNB2; LRRK2; MAN2B1; MAPT; MECP2; MLC1; MMADHC; MOCS1; MR1; MRE11A; MSTO1; MTPAP; MTTP; NALCN; NDUFA10; NDUFA12; NDUFA2; NDUFA9; NDUFAF2; NDUFAF6; NDUFS3; NDUFS4; NDUFS7; NDUFS8; NEU1; NFASC; NHLRC1; NKX2-1; NKX6-2; NOL3; NOP56; NPC1; NPC2; NPHP1; NR4A2; NTN1; NUS1; OFD1; OPA1; OPA3; OPHN1; PAH; PARK2; PARK7; PAX6; PCBD1; PCCA; PCCB; PDE10A; PDE2A; PDGFB; PDGFRB; PDHA1; PDHX; PDSS1; PDSS2; PDYN; PEX10; PEX2; PEX7; PHYH; PI4K2A; PIK3R5; PINK1; PITRM1; PLA2G6; PLEKHG4; PLP1; PLXNB3; PMM2; PMPCA; PMPCB; PNKD; PNKP; PNPLA6; PODXL; POLG; POLR1C; POLR3A; POLR3B; PPP2R2B; PPP2R5D; PRICKLE1; PRKCG; PRKRA; PRNP; PRPS1; PRRT2; PTRH2; PTS; QDPR; RAB39B; RAB7L1; RAD51; RARS2; REEP1; RELN; RIC3; RNF168; RNF170; RNF216; RPGRIP1L; SACS; SAMD9L; SCARB2; SCN1A; SCN2A; SCN8A; SCP2; SCYL1; SDHA; SDHAF1; SEPSECS; SERAC1; SETD5; SETX; SGCE; SIL1; SLC16A2; SLC17A5; SLC18A2; SLC19A3; SLC1A3; SLC20A2; SLC23A2; SLC2A1; SLC30A10; SLC30A2; SLC39A14; SLC46A1; SLC52A2; SLC6A19; SLC6A3; SLC6A5; SLC9A1; SNCA; SNX14; SOX6; SPAST; SPG11; SPG7; SPR; SPTBN2; SQSTM1; STUB1; SUCLA2; SUCLG1; SUOX; SURF1; SYNE1; SYNJ1; SYT1; SYT14; TACF1; TBC1D24; TBP; TDP1; TDP2; TGM6; TH; THAP1; THG1L; TIMM8A; TMEM216; TMEM230; TMEM240; TMEM67; TOR1A; TPP1; TRIM17; TRNT1; TSEN2; TSEN34; TSEN54; TTBK2; TTC19; TTPA; TUBB4A; UBA5; UCHL1; UNC13A; UQCRQ; VAC14; VAMP1; VAMP2; VLDLR; VPS13A; VPS13D; VPS35; VWA3B; WDR45; WDR73; WDR81; WFS1; WWOX; XK; XPA; XPC; XPR1; ZFYVE26; ZNF 142; ZNF574; ZNF592

Targeted gene panel sequencing and whole Exome Sequencing

Genomic DNA from peripheral blood samples was extracted according to standard procedures. An exome/targeted gene capture kit (SeqCap EZ Exome probes; Roche-NimbleGen) was used to target all exons. Exons/genes capture was followed by massive parallel 150pb paired-end sequencing (Illumina, San Diego, CA, USA). Read mapping and variant calling were performed following standard bioinformatics procedures [1]. Filtering and prioritization of the variants were conducted using an in-house interactive Paris Descartes bioinformatics platform pipeline based on the Ensembl database (release 67). Variants were filtered according to their frequency (1%) against the dbSNP (https://www.ncbi.nlm.nih.gov/projects/SNP/), 1000 Genome Project (http://www.internationalgenome.org/), ExAC (http://exac.broadinstitute.org/) and GnomAD (http://gnomad.broadinstitute.org/about) databases. In silico prediction of variants pathogenicity was performed using SIFT, PolyPhen and Mutation Taster. CADD score was also integrated to the prioritization criteria, and a cutoff of 20 was used to determine deleteriousness. Pathogenic short tandem repeats in coding sequence was performed through the use of the bioinformatics tool ExpansionHunter [2].

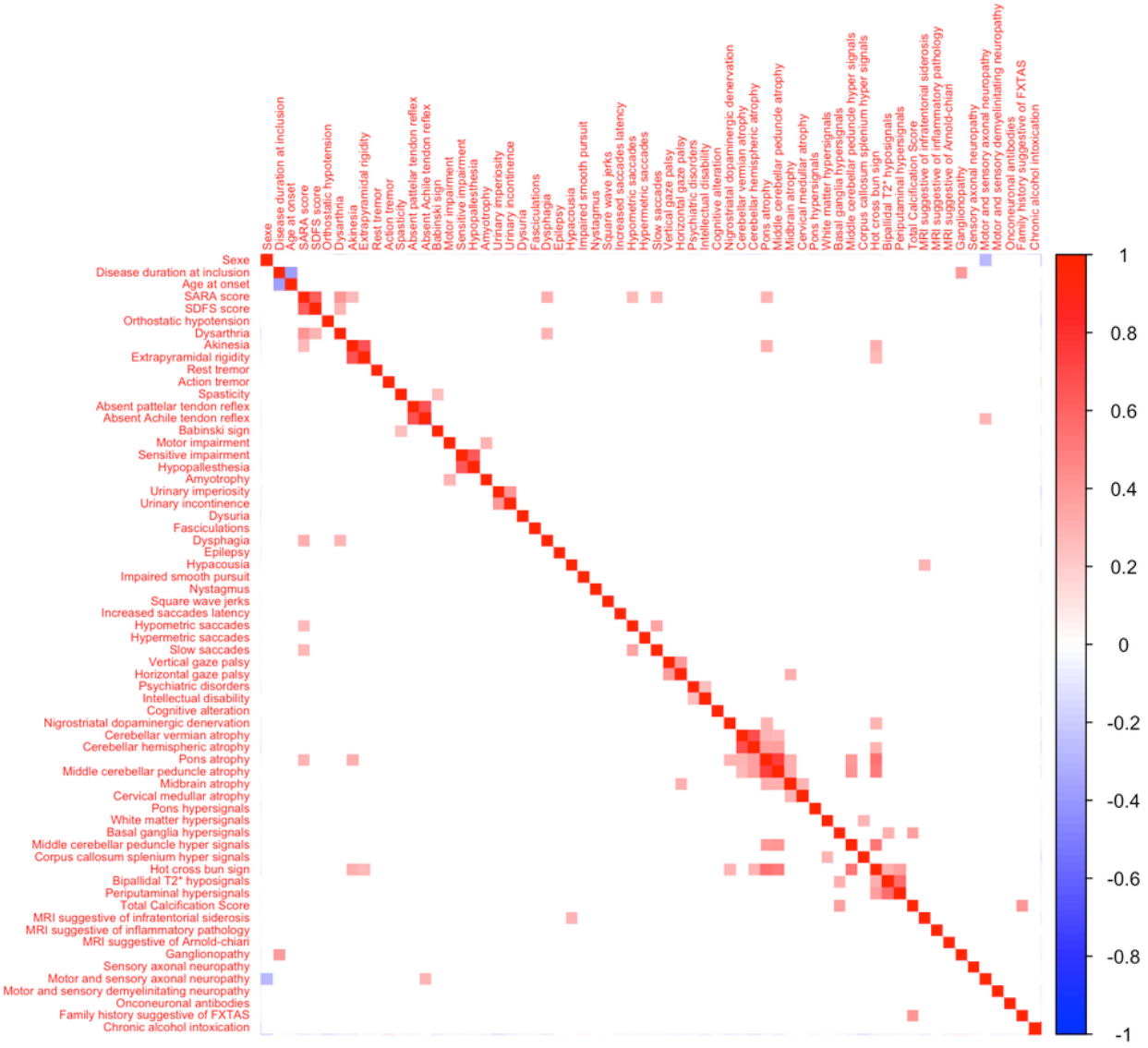
Circular correlation plot and correlation matrices

1) *Circular correlation plot with only significant associations*



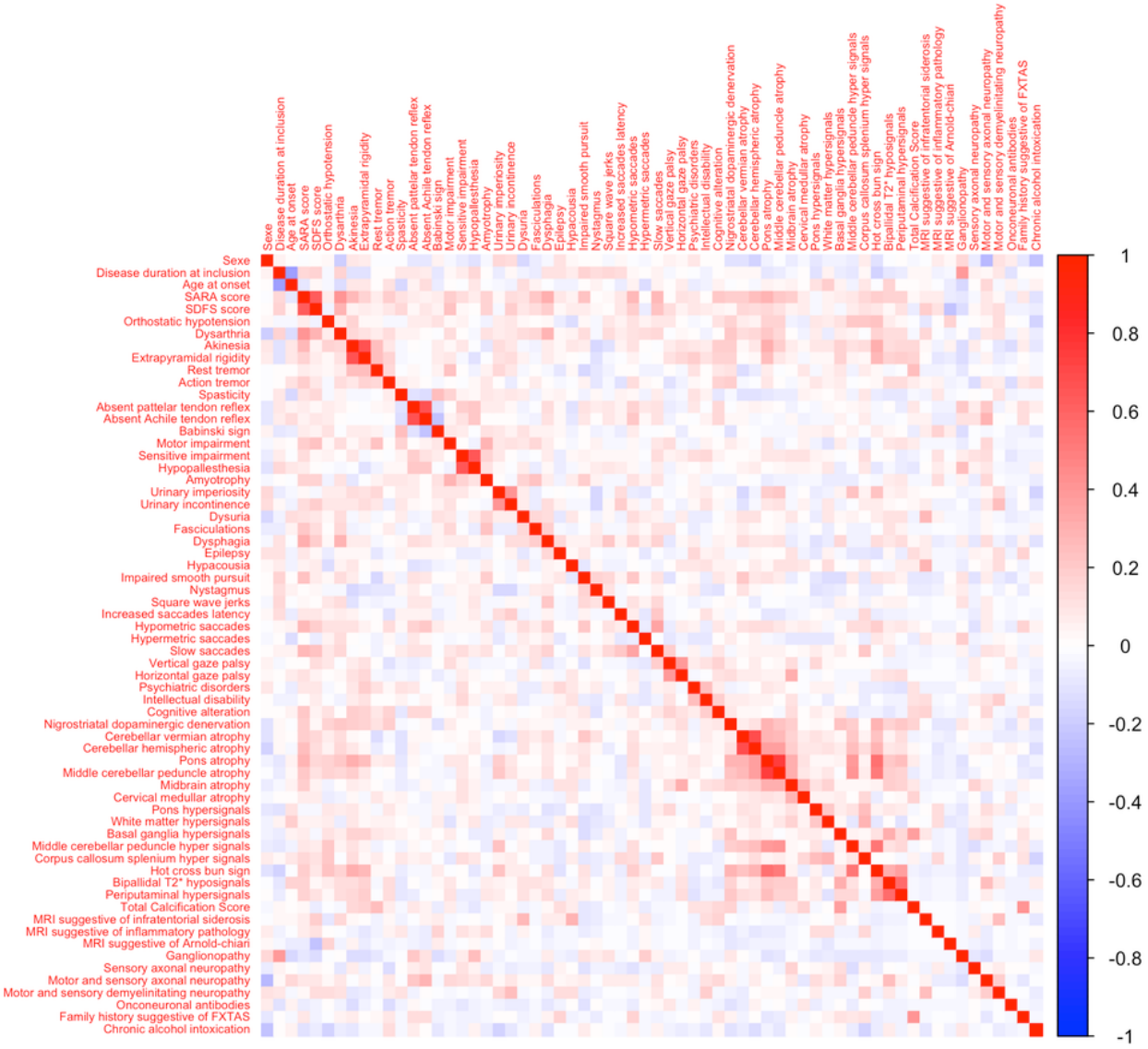
Circular correlation plot showing correlation between two variables.
Only significant associations with p<0,05 is represented.

2) *Correlation matrix with only significant associations*



Correlation matrix plot showing, in a different representation, correlation between two variables. The variables are listed in the x and y-axis. Each cell in the matrix shows the strength (stronger if darker) and direction (red positive, blue negative) of the association. This plot shows only significant association with p<0,05.

3) Correlation matrix with all associations



This plot shows all the association, regardless of their significance.

Multiple Factor Analysis variable groups

Group 1: Sexe

Group 2: Disease duration a inclusion; Age of ataxia onset

Group 3: Akinesia; Rigidity; Rest Tremor; Action Tremor

Group 4: Spasticity; Babinski sign; Motor deficit

Group 5: Absent patellar tendon reflex; Absent Achille tendon reflex; fasciculations; Sensitive deficit; Hypopallesthesia; Amyotrophy

Group 6: Orthostatic hypotension; Urinary imperiosity; Urinary incontinence; Dysuria;

Group 7: Dysarthria; Dysphagia

Group 8: Epilepsy; Family history suggestive of FXTAS; Chronic alcohol intoxication; Intellectual disability; Psychiatric disorders; Cognitive alterations; Hypoacousia

Group 9: Impaired smooth pursuit; Nystagmus; Square wave Jerks; Increased saccades latencies; Hypometric saccades; Hypermetric saccades; Slow saccades; Vertical gaze palsy; Horizontal gaze palsy

Group 10: Nigrostriatal dopaminergic denervation

Group 11: Cerebellar vermian atrophy; Cerebellar hemispheric atrophy; Pons atrophy; Middle cerebellar peduncle atrophy; Midbrain atrophy; Cervical medullar atrophy; Pons hypersignals; White matter hypersignals; Basal ganglia hypersignals; Middle cerebellar peduncle hyper signals; Corpus callosum splenium hypersignals; Hot cross bun sign; Bipallidal T2* hyposignals; Periputaminal rim hypersignals; MRI suggestive of non-genetic causes (multiple sclerosis, Arnold-Chiari malformations, Creutzfeld Jakob disease, superficial siderosis, fossa posterior tumor or pons cavernoma); Pathological Total Calcification Score

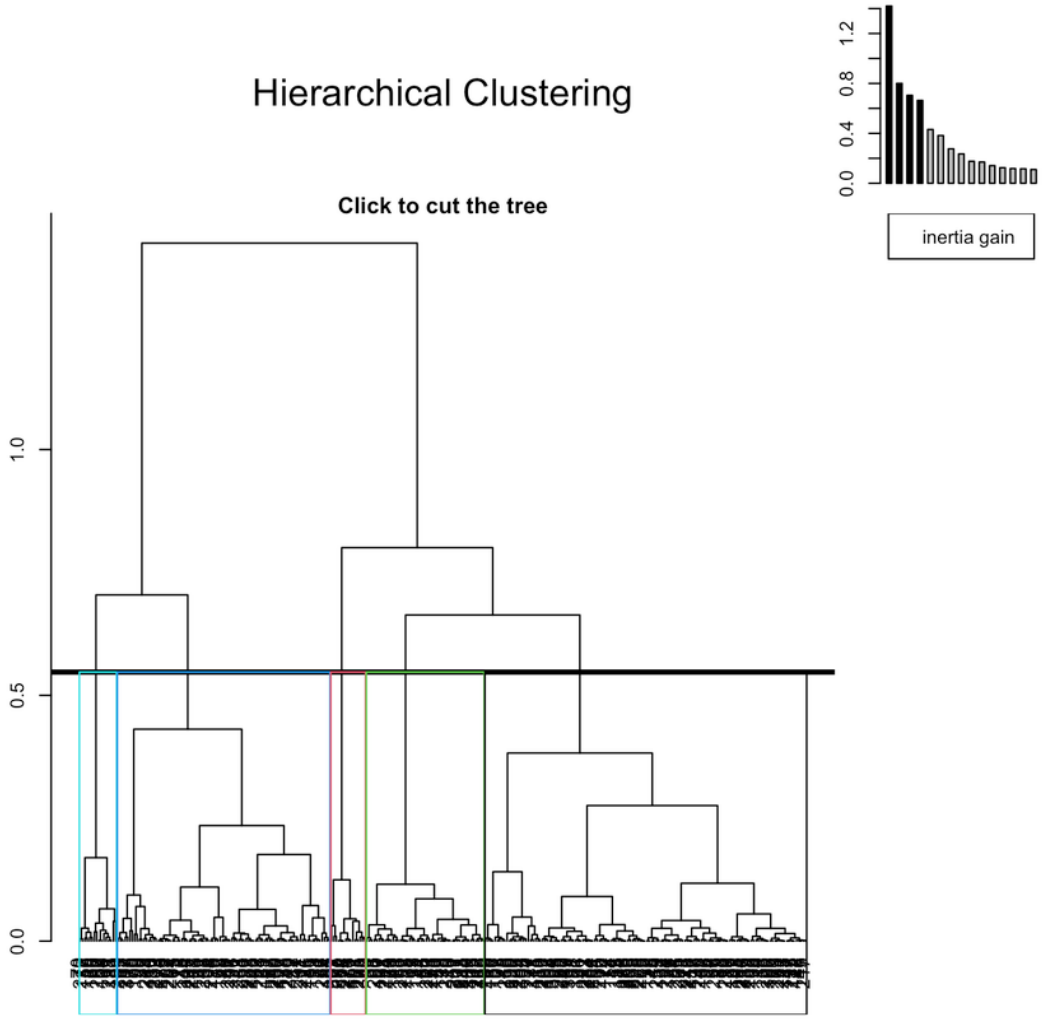
Group 12: Sensitive neuronopathy; Sensory axonal neuropathy; Motor and sensory axonal neuropathy; motor and sensory demyelinating neuropathy

Group 13: Onconeuronal antibodies

Group 14: SARA score; SARA/Disease duration; SDFS score; SDFS/Disease duration

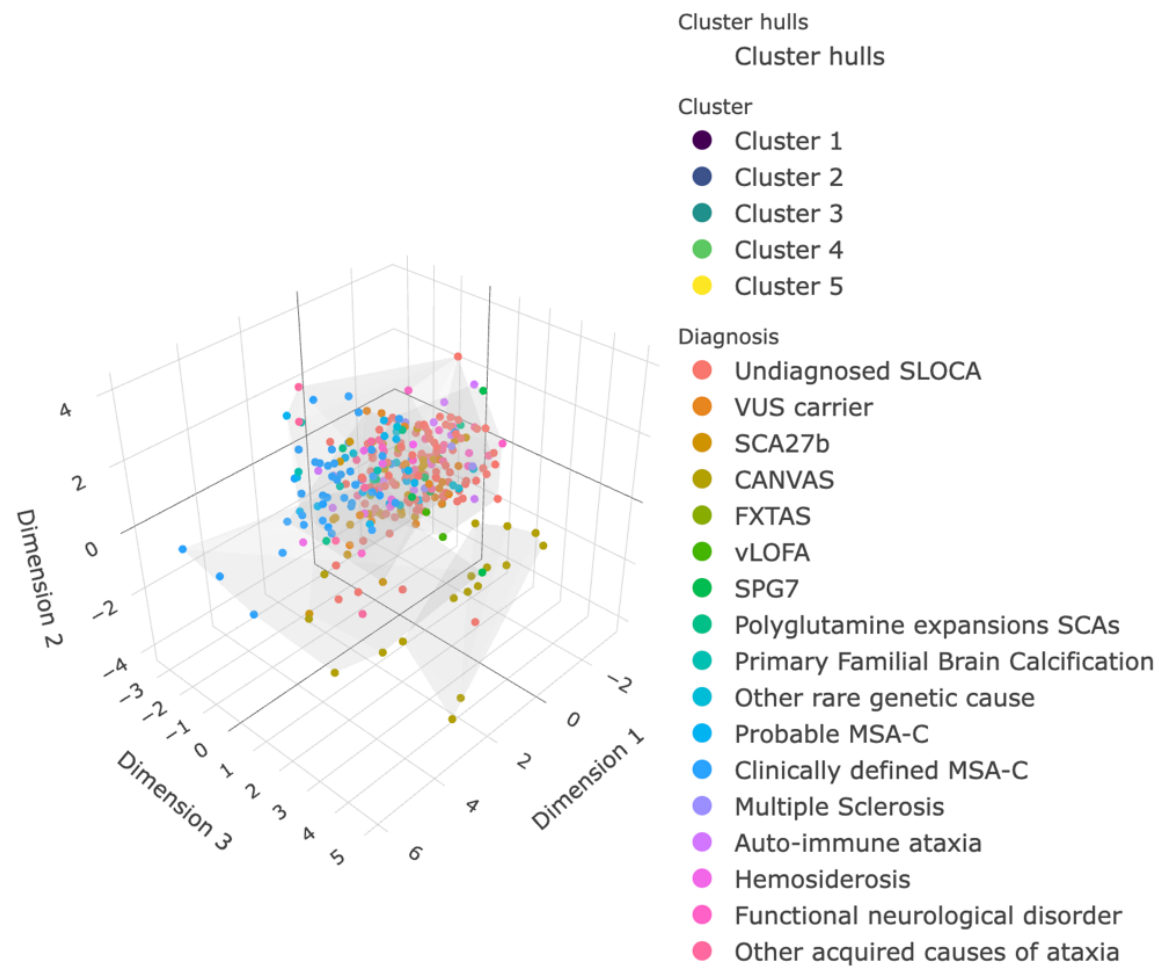
Group 15: Diagnosis (supplementary variable not used for principal components extraction

Cluster dendogram



Dendograph visualizing in a tree-like diagram the arrangement of patients into clusters formed by hierarchical clustering. In the horizontal axis with each line at the bottom are representing a patient. The branches above them represent clusters of several patients formed at each step for the hierarchical clustering processes. Patients connected by short branches (bottom of the figure) are more similar to each other's, while those connected by a branche on the top are more dissimilar and therefore grouped in a cluster at a higher level of the hierarchical clustering. The horizontal line across the dendogram shows the final clusters, represented in color. The inertia gain in the top right helps to deciding the optimal number of clusters (where the inertia drop), by measuring how tightly the patients are clustered together. 4 to 5 clusters seems the be the right number of clusters.

Principal components tridimensional plot with clusters highlighted

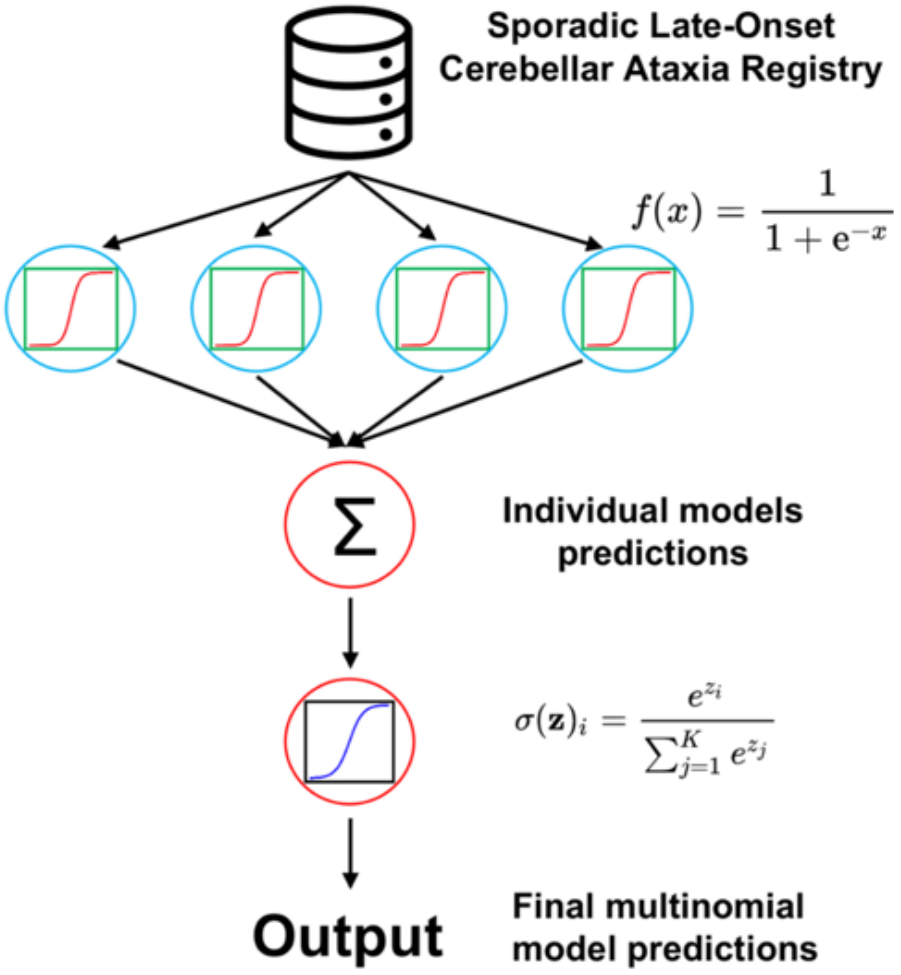


3D representation of the multiple factor analysis. It shows in a 3D space the distribution of patients across the three principal components derived from the Multiple Factor Analysis. It represents how patients (points) color-coded by their diagnosis are clustered together.

Table of correlation between clusters and causal diagnoses.
Number and proportion of patients in each clusters according to their diagnosis.

	CLUSTER 1	CLUSTER 2	CLUSTER 3	CLUSTER 4	CLUSTER 5
ILOCA, N (%)	49 (51.57)	27 (28.42)	1 (1.05)	14 (14.73)	4 (4.21)
VUS, N (%)	10 (52.63)	3 (15.78)	0 (0.00)	3 (15.78)	3 (15.78)
SCA27B, N (%)	8 (30.76)	14 (53.84)	0 (0.00)	3 (11.53)	1 (3.84)
CANVAS, N (%)	1 (5.55)	0 (0.00)	13 (72.22)	0 (0.00)	4 (22.22)
FXTAS, N (%)	0 (0.00)	3 (100.00)	0 (0.00)	0 (0.00)	0 (0.00)
VLOFA, N (%)	1 (50.00)	0 (0.00)	0 (0.00)	1 (50.00)	0 (0.00)
SPG7, N (%)	4 (57.14)	1 (14.28)	0 (0.00)	1 (14.28)	1 (14.28)
SCA POLYQ, N (%)	1 (16.66)	1 (16.66)	0 (0.00)	3 (50.00)	1 (16.66)
PFBC, N (%)	0 (0.00)	0 (0.00)	0 (0.00)	4 (100.00)	0 (0.00)
OTHER RARE GENETIC CAUSE, N (%)	3 (33.33)	5 (55.55)	0 (0.00)	1 (11.11)	0 (0.00)
PROBABLE MSA-C, N (%)	3 (18.75)	1 (6.25)	0 (0.00)	10 (62.50)	2 (12.50)
CLINICALLY DEFINED MSA-C, N (%)	5 (10.63)	2 (4.25)	0 (0.00)	31 (65.95)	9 (19.14)
MULTIPLE SCLEROSIS, N (%)	3 (60.00)	2 (40.00)	0 (0.00)	0 (0.00)	0 (0.00)
AUTO-IMMUNE ATAXIA, N (%)	7 (63.63)	2 (18.18)	0 (0.00)	2 (18.18)	0 (0.00)
HEMOSIDEROSIS, N (%)	4 (57.14)	0 (0.00)	0 (0.00)	2 (28.57)	1 (14.28)
FUNCTIONAL NEUROLOGICAL DISORDER, N (%)	6 (50.00)	2 (16.66)	0 (0.00)	4 (33.33)	0 (0.00)
OTHER ACQUIRED CAUSES OF ATAXIA, N (%)	10 (50.00)	3 (15.00)	0 (0.00)	4 (20.00)	3 (15.00)

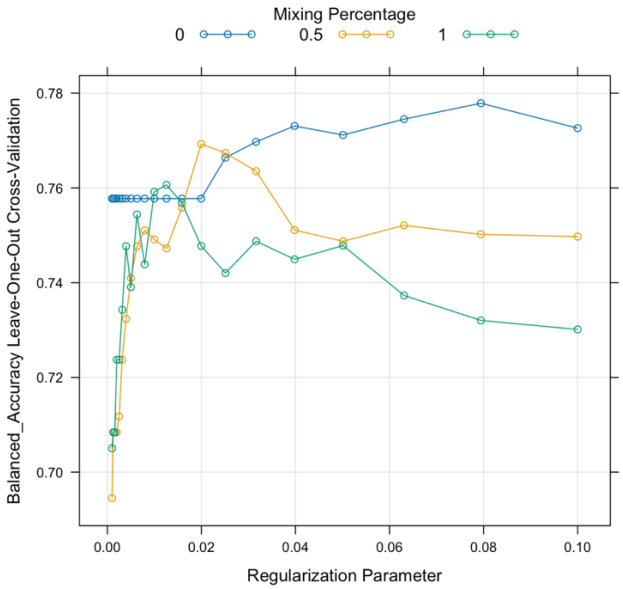
Binary relevance ensemble model



Four logistic regression models were trained to predict non-genetic causes of SLOCA, ATX-FGF14b/SCA27b, ATX-RFC1/CANVAS, rare genetic causes of SLOCA or ILOCA. The predictions of those models were then extracted to train a final multinomial model based on the Softmax activation function.

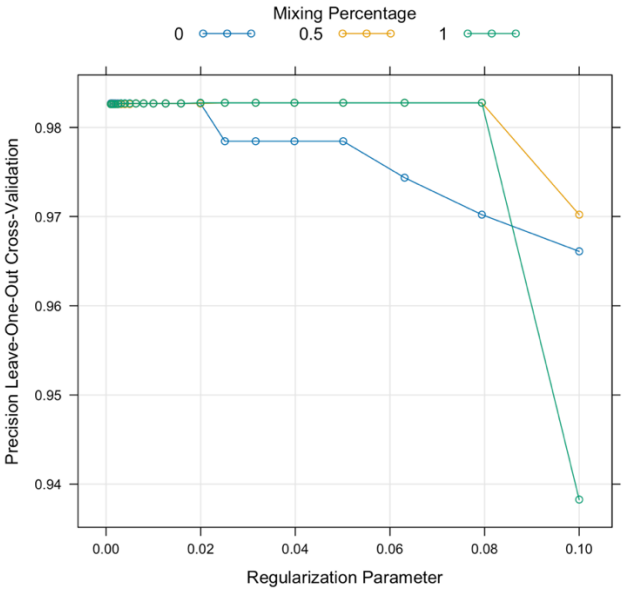
Hyperparameter tuning plots

1) *Hyperparameter tuning of the non-genetic binary model*



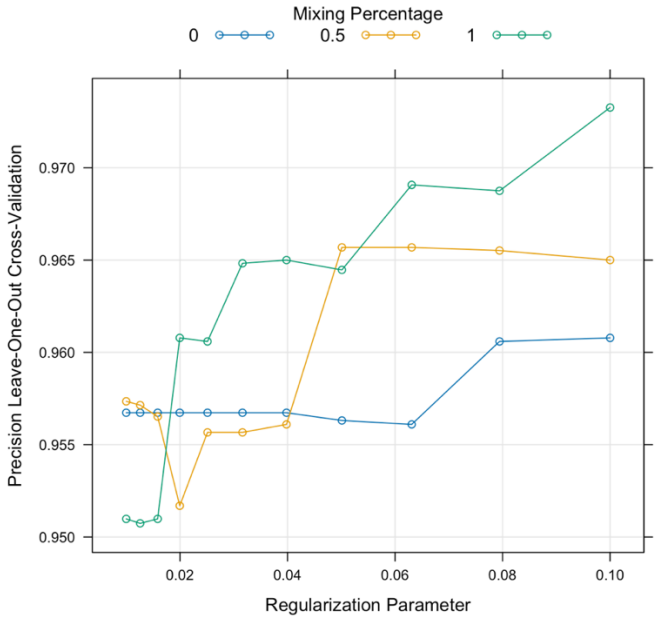
This figure shows the performance (balanced accuracy) of the non-genetic binary model across different regularization parameters (from 0 to 0,1) using to penalized variables from the model in the aim to make a model neither underfitted nor overfitted. The x-axis represents the regularization parameter values and the y-axis the balanced accuracy of this model. The regularization parameter adjusts the penalty applied to the coefficient of each variable t of the model (higher penalization leading to a simpler model with less overfitting). The curve represents different mixing percentage of the penalization: alpha=0 for the blue curve (L2 ridge pure), alpha=1 for the green curve (L1 Lasso pure) et alpha= 0,5 for the orange curve is a mix between those 2 types of penalization (Elastic net).

2) *Hyperparameter tuning of the ATX-RFC1/CANVAS binary model*



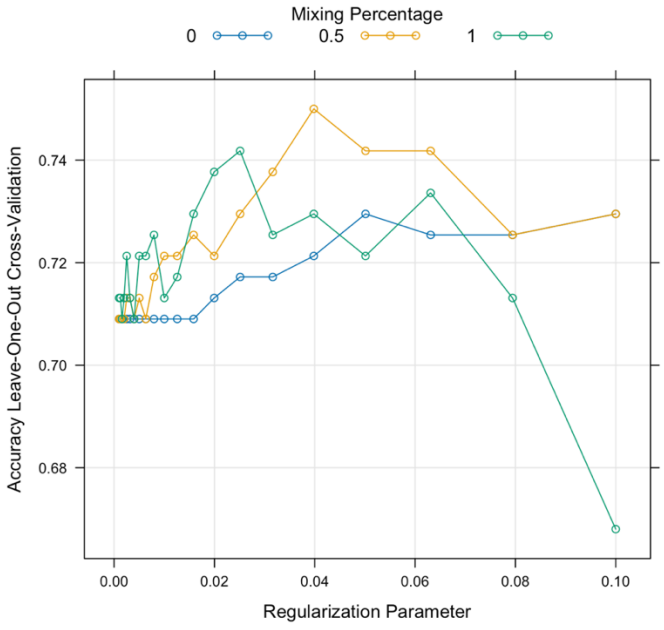
This figure shows the performance (balanced accuracy) of the CANVAS binary model across different regularization parameters (from 0 to 0,1) using to penalized variables from the model in the aim to make a model neither underfitted nor overfitted.

3) Hyperparameter tuning of the ATX-RFC1/SCA27b binary model



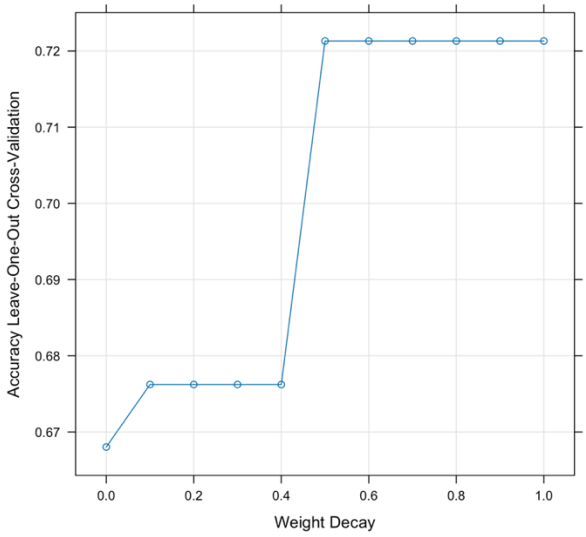
This figure shows the performance (balanced accuracy) of the SCA27b binary model across different regularization parameters (from 0 to 0,1) using to penalized variables from the model in the aim to make a model neither underfitted nor overfitted.

4) Hyperparameter tuning of the rare genetic or ILOCA binary model



This figure shows the performance (balanced accuracy) of the rare genetic or ILOCA binary model across different regularization parameters (from 0 to 0,1) using to penalized variables from the model in the aim to make a model neither underfitted nor overfitted.

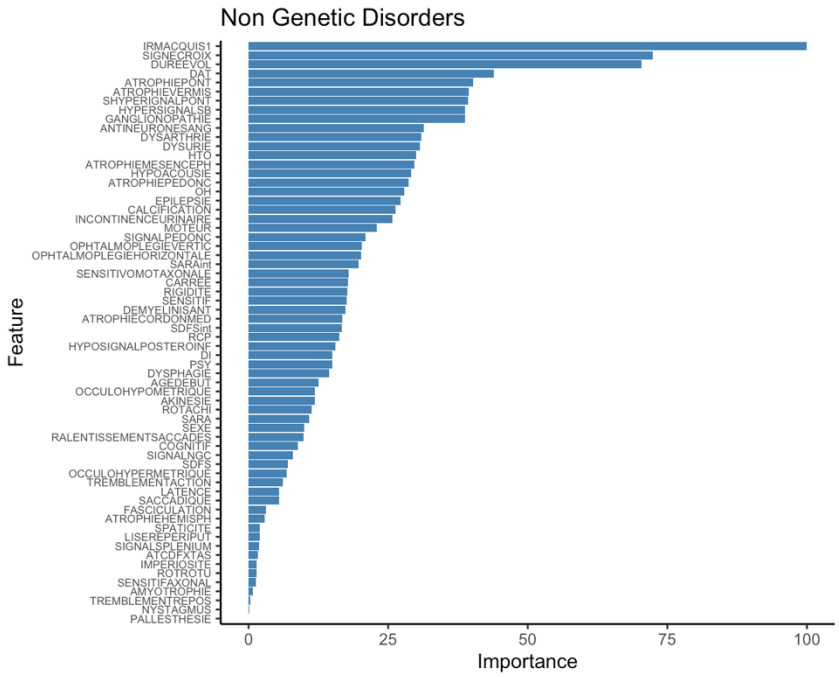
5) Hyperparameter tuning of the ensemble/multinomial model



This figure shows the performance (balanced accuracy) of the multinomial model across different regularization parameters (from 0 to 1). The x-axis represents the weight decay which is a regularization technique to prevent overfitting. The y-axis shows the model's accuracy score. At a weight decay value around 0,5, there is a significant jump in the performance of the model with an accuracy of 0,72 that stabilize after, meaning the regularization improves the model's performance. At this weight decay, the model is parsimonious, avoiding overfitting and still capturing the important pattern of the dataset.

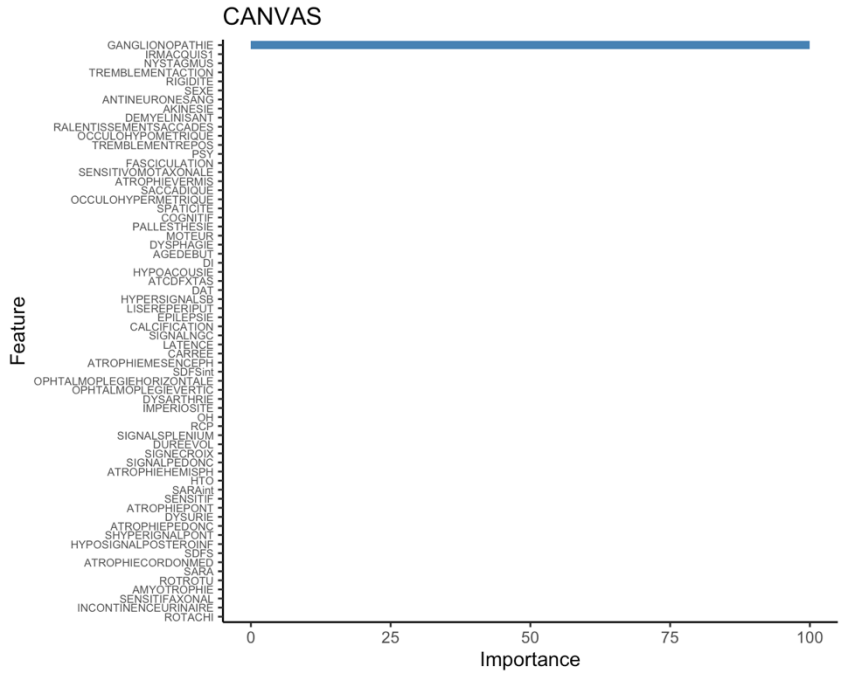
Importance of variables for binary models

1) *Importance of variables for non-genetic binary model*



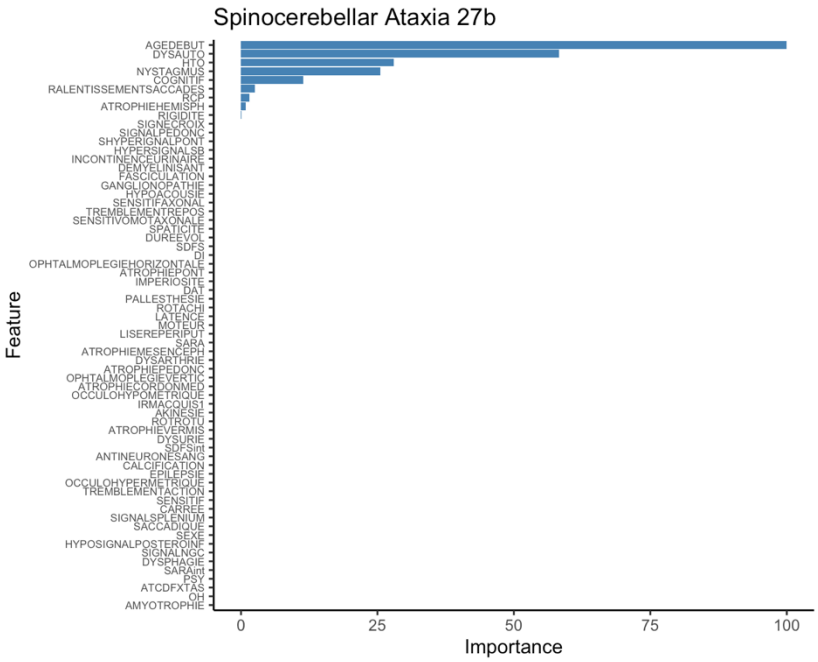
This figure shows the importance (represented by the length of the horizontal bar) for each feature retained in the non-genetic binary model.

2) *Importance of variables for CANVAS binary model*



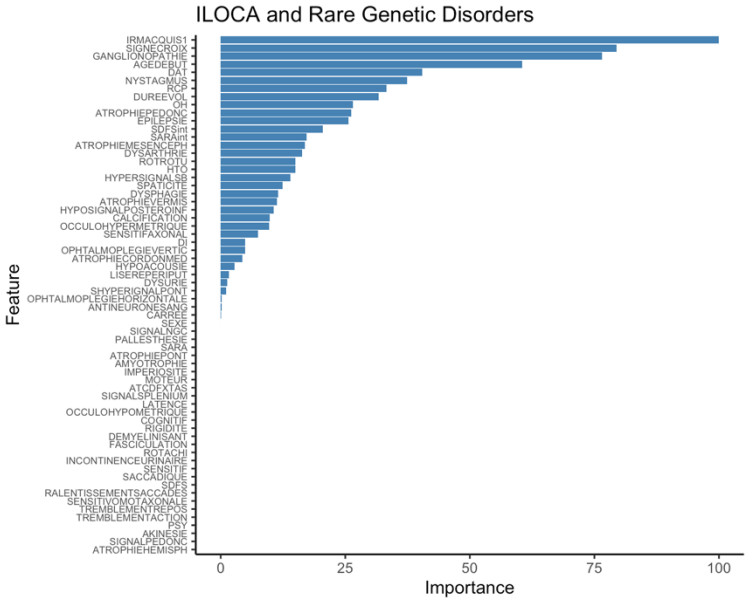
This figure shows the importance (represented by the length of the horizontal bar) for each feature retained in the CANVAS model.

3) *Importance of variables for SCA27b binary model*



This figure shows the importance (represented by the length of the horizontal bar) for each feature retained in the SCA27b binary model.

4) *Importance of variables rare genetic disorders and ILOCA binary model*



This figure shows the importance (represented by the length of the horizontal bar) for each feature retained in the rare genetic or ILOCA binary model.

Model coefficients

1) Binary non genetic model: coefficient of each variable retained in the model

RETAINED VARIABLE	COEFFICIENT
Sexe	0.05803063172174331
Disease duration at inclusion	-0.4063517699364935
Age at onset	0.07290628751800263
SARA	0.06327190050495095
SARAint	0.1145341583424925
SDFS	0.04128625716601553
SDFSint	0.09682403199269878
Orthostatic hypotension	0.1734857298396012
Dysathria	0.1786203419305998
Akinesia	0.06908692953661565
Extrapyramidal rigidity	0.1025759459738794
Rest tremor	-0.002359155886656186
Action tremor	-0.0358418576323373
Spasticity	-0.01224557889758695
Pattelar tendon reflex	-0.008696312434370967
Achileen tendon reflex	-0.06560411193771573
Babinski sign	-0.09392255602493656
Motor impairment	-0.1331768795053187
Sensory impairment	-0.1020888860903994
Pallesthsia	-0.0007145704144962579
Amyotrophia	0.005011966779532289
Urinary imperiosity	-0.008758570815792722
Urinary incontinence	0.1489095658992097
Dysuria	0.1772483988559386
Fasciculation	0.01858403375758642
Dysphagia	0.08397497885401813
Epilepsy	-0.1575081146756624
Hypoacousia	0.1683926060784245
Impaired smooth poursuite	0.0319630306062649
Nystamgus	-0.001047099736409693
Square wave jerks	-0.1033120382582541
Saccades latency	0.03246906196886605
Hypometric saccades	-0.06914770028329652
Hypermetric saccades	-0.04000462517649362
Slow saccades	-0.05769087545417926
Verticale gaze palsy	0.1171779957552354
Horizontale gaze palsy	0.1167285531596679
Psychiatric disorders	-0.08707556296511212
Mental dissability	-0.08740750223526192
Altered neuropsychological assement	0.05187581672605434
Nigrostriatal dopaminergic denervation	0.2539094300539351
MRI suggestive of non-generative acquired cause	0.5769217478714975
MRI suggestive of non-generative acquired cause	-0.2279070269338919
Cerebellar hemispheric atrophy	-0.01702427834260505
Pons atrophy	0.2324895170408395
Middle Cerebellar peduncle atrophy	0.1657611929306021
Midbrain atrophy	0.1714542208093613
Cervical medullar atrophy	-0.09774621552383518
Pons hypersignals	0.2273536377215651
Hemispheric white matters hypersignals	-0.2240475999444407
Basal ganglia hypersignal	0.04647026362498457
Middle Cerebellar peduncle	0.121089969635714
Splenium hypersignal	-0.01137792110960364
Hot cross bun sign	0.4180563552738349
Postero-inferior hyposignal	0.09026775829955755
Bipallidal T2* hyposignal	0.01222303630341768
Basal ganglia calcifications	0.1521492180376678
Ganglionopathy	-0.2237364910695882
Sensory axonal neuropathy	-0.008273554556430415
Motor and sensory axonal neuropathy	-0.1039823237341323
Motor and sensory demyelination neuropathy	0.100943887659605
Positive onconeuronral antibodies	0.1814179943726192
Family history of FXTAS	0.01053814411982903
Chronic Alcohol intoxication	-0.1615566074053451

2) Binary ATX-RFC1/CANVAS model: coefficient of each variable retained in the model

RETAINED VARIABLE	COEFFICIENT
Ganglionopathy	0.8126507475837327

3) Binary ATX-FGF14 model: coefficient of each variable retained in the model

RETAINED VARIABLE	COEFFICIENT
Age at onset	1.060359343598827
Orthostatic hypotension	-0.2963517852986783
Extrapyramidal rigidity	-0.001316916706691007
Babinski sign	-0.01622850712735228
Dysautonomia	-0.617548177485365
Nystagmus	0.2707430463120322
Slow saccades	-0.02743998821375222
Altered neuropsychological assessment	-0.1207147600648857
Cerebellar hemispheric atrophy	-0.008756416568500217

4) Binary rare genetic disorders or ILOCA: coefficient of each variable retained in the model

RETAINED VARIABLE	COEFFICIENT
Sexe	-0.000008283902835038865
Disease duration at inclusion	0.1998888425362517
Age at onset	-0.3812381628218697
SARA/disease duration	-0.108752172891497
SDFS/disease duration	-0.1290986844181234
Orthostatic hypotension	-0.09440019430316957
Dysathria	-0.1028572654697047
Spasticity	0.07815037324262003
Pattelar tendon reflex	-0.09478046162176429
Babinski sign	0.2099606025861777
Dysuria	-0.008053848739770319
Dysphagia	-0.0725811904839115
Epilepsy	0.1615728470525667
Hypoacousia	-0.01756568663532434
Nystamgus	-0.2360261947851971
Square wave jerks	-0.0008084078777921024
Hypermetric saccades	-0.06127924122310378
Verticale gaze palsy	-0.03078806470757677
Horizontale gaze palsy	-0.001464942826757889
Mental disability	0.03093292571069096
Nigrostriatal dopaminergic denervation	-0.255384778889957
MRI suggestive of non-generative acquired cause	-0.6306655776839398
Cerebellar vermian atrophy	0.07153178355968177
Middle Cerebellar peduncle atrophy	-0.1654904249577815
Midbrain atrophy	-0.10672374263901
Cervical medullar atrophy	0.02727113694985465
Pons hypersignals	-0.006798742069347759
Hemispheric white matters hypersignals	0.08812011986061996
Hot cross bun sign	-0.5008918939443635
Postero-inferior hyposignal	-0.06708748513988617
Bipallidal T2* hyposignal	-0.01036876809906235
Basal ganglia calcifications	-0.06188026905735637
Ganglionopathy	-0.4824714648318414
Sensory axonal neuropathy	0.04705909361834877
Positive onconeuronal antibodies	-0.0009803020381562416
Chronic Alcohol intoxication	0.1673940890358461

5) Ensemble model: Coefficient of each binary model in the ensemble model to predict either a CANVAS, a SCA27b or a rare genetic disorders/ILOCA

	NON GENETIC BINARY MODEL	SCA27B BINARY MODEL	CANVAS BINARY MODEL	RARE GENETIC DISORDERS/ILOCA BINARY MODEL
ATX-RFC1/CANVAS	-0.4634374448647286	-0.1210275777490957	0.8448668459028797	0.03618955343213971
ATX-FGF14/SCA27b	-0.4740195733264816	0.633162984789945	0.04772012662517766	0.1699136773287171
RARE GENETIC DISORDERS/ILOCA	-0.9149253525346848	-0.1273589452152785	-0.04401315171152372	0.5334898332761602

Detailed model performances

1) Multinomial model

	Accuracy	Kappa	No information rate	P-Value [Acc > NIR]	logLoss
Internal validation dataset					
Multinomial model	**72%**	0	/	/	0.79
External validation dataset					
Multinomial model	**71%**	0.51	0.47	0.00019	/

2) Binary models

	Accuracy	Sensitivity	Specificity	Positive Predictive Value	Negative Predictive Value	Prevalence	Detection rate	Detection prevalence
Internal validation dataset								
Binary model Non genetic	**78%**	66%	89%	81%	81%	39%	26%	32%
Binary model ATX-RFC1	**86%**	73%	99%	91%	98%	6%	4%	5%
Binary model ATX-FGF14	**72%**	52%	91%	38%	95%	8%	4%	12%
Binary model rare genetic cause/ILOCA	**77%**	80%	74%	73%	81%	46%	37%	51%
External validation dataset								
Binary model Non genetic	**76%**	65%	86%	75%	79%	38%	25%	34%
Binary model ATX-RFC1	**100%**	100%	100%	100%	100%	5%	5%	5%
Binary model ATX-FGF14	**69%**	40%	98%	67%	95%	8%	3%	5%
Binary model rare genetic cause/ILOCA	**71%**	78%	65%	67%	77%	47%	37%	56%

Misclassified patients and their respective diagnosis

1) Internal validation

ID	MODEL PREDICTION	ACCURATE PREDICTION	DIAGNOSIS DETAILS
STARAC1	NON GENETIC	ATX-FGF14/SCA27b	ATX-FGF14/SCA27b
STARAC12	NON GENETIC	RARE GENETIC DISEASE OR ILOCA	VUS carrier
STARAC15	NON GENETIC	RARE GENETIC DISEASE OR ILOCA	ILOCA
STARAC19	ATX-FGF14/SCA27b	RARE GENETIC DISEASE OR ILOCA	ILOCA
STARAC25	ATX-FGF14/SCA27b	RARE GENETIC DISEASE OR ILOCA	ILOCA
STARAC33	NON GENETIC	RARE GENETIC DISEASE OR ILOCA	ATX-ATXN3/SCA3
STARAC39	ATX-FGF14/SCA27b	RARE GENETIC DISEASE OR ILOCA	ILOCA
STARAC43	ATX-FGF14/SCA27b	RARE GENETIC DISEASE OR ILOCA	ATX-PRKCG/SCA14
STARAC44	ATX-FGF14/SCA27b	RARE GENETIC DISEASE OR ILOCA	ILOCA
STARAC49	ATX-FGF14/SCA27b	NON GENETIC	Clinically probable MSA
STARAC51	RARE GENETIC DISEASE OR ILOCA	NON GENETIC	Clinically defined MSA
STARAC56	RARE GENETIC DISEASE OR ILOCA	ATX-RFC1/CANVAS	ATX-RFC1/CANVAS
STARAC57	RARE GENETIC DISEASE OR ILOCA	NON GENETIC	Autoimmune encephalitis
STARAC61	ATX-RFC1/CANVAS	RARE GENETIC DISEASE OR ILOCA	ILOCA
STARAC63	ATX-FGF14/SCA27b	RARE GENETIC DISEASE OR ILOCA	VUS carrier
STARAC67	RARE GENETIC DISEASE OR ILOCA	NON GENETIC	Functional neurodisorders
STARAC70	ATX-FGF14/SCA27b	NON GENETIC	Gayet Wernicke encephalopathy
STARAC72	RARE GENETIC DISEASE OR ILOCA	NON GENETIC	Clinically defined MSA
STARAC73	NON GENETIC	RARE GENETIC DISEASE OR ILOCA	ATX-ATXN2/SCA2
STARAC77	NON GENETIC	RARE GENETIC DISEASE OR ILOCA	ILOCA
STARAC78	RARE GENETIC DISEASE OR ILOCA	NON GENETIC	Post-traumatic ataxia
STARAC79	RARE GENETIC DISEASE OR ILOCA	NON GENETIC	Clinically defined MSA
STARAC89	RARE GENETIC DISEASE OR ILOCA	NON GENETIC	Sjögren syndrome
STARAC91	ATX-FGF14/SCA27b	NON GENETIC	Post-infection ataxia
STARAC98	NON GENETIC	RARE GENETIC DISEASE OR ILOCA	VUS carrier
STARAC108	RARE GENETIC DISEASE OR ILOCA	NON GENETIC	Clinically probable MSA
STARAC116	RARE GENETIC DISEASE OR ILOCA	ATX-FGF14/SCA27b	ATX-FGF14b/SCA27b
STARAC121	RARE GENETIC DISEASE OR ILOCA	NON GENETIC	Functional neurodisorders
STARAC129	RARE GENETIC DISEASE OR ILOCA	ATX-FGF14/SCA27b	ATX-FGF14/SCA27b
STARAC131	ATX-FGF14/SCA27b	RARE GENETIC DISEASE OR ILOCA	ILOCA
STARAC135	NON GENETIC	ATX-FGF14/SCA27b	ATX-FGF14/SCA27b
STARAC136	RARE GENETIC DISEASE OR ILOCA	NON GENETIC	Clinically probable MSA
STARAC137	RARE GENETIC DISEASE OR ILOCA	ATX-RFC1/CANVAS	ATX-RFC1/CANVAS
STARAC140	ATX-FGF14/SCA27b	NON GENETIC	Gougerot-Sjögren syndrome
STARAC142	ATX-FGF14/SCA27b	NON GENETIC	Functional neurodisorders
STARAC147	RARE GENETIC DISEASE OR ILOCA	ATX-FGF14/SCA27b	ATX-FGF14/SCA27b
STARAC152	ATX-FGF14/SCA27b	RARE GENETIC DISEASE OR ILOCA	ILOCA
STARAC178	RARE GENETIC DISEASE OR ILOCA	NON GENETIC	Clinically probable MSA
STARAC186	RARE GENETIC DISEASE OR ILOCA	NON GENETIC	Functional neurodisorders
STARAC188	RARE GENETIC DISEASE OR ILOCA	NON GENETIC	Functional neurodisorders
STARAC191	RARE GENETIC DISEASE OR ILOCA	NON GENETIC	Vascular
STARAC201	RARE GENETIC DISEASE OR ILOCA	NON GENETIC	Clinically probable MSA
STARAC203	NON GENETIC	ATX-FGF14/SCA27b	ATX-FGF14/SCA27b
STARAC204	RARE GENETIC DISEASE OR ILOCA	NON GENETIC	Vascular
STARAC205	ATX-FGF14/SCA27b	NON GENETIC	Autoimmune encephalitis
STARAC206	NON GENETIC	RARE GENETIC DISEASE OR ILOCA	ILOCA
STARAC207	RARE GENETIC DISEASE OR ILOCA	ATX-FGF14/SCA27b	ATX-FGF14/SCA27b
STARAC208	NON GENETIC	ATX-RFC1/CANVAS	ATX-RFC1/CANVAS
STARAC213	RARE GENETIC DISEASE OR ILOCA	ATX-FGF14/SCA27b	ATX-FGF14/SCA27b
STARAC215	RARE GENETIC DISEASE OR ILOCA	NON GENETIC	Functional neurodisorders
STARAC216	NON GENETIC	RARE GENETIC DISEASE OR ILOCA	VUS carrier
STARAC217	RARE GENETIC DISEASE OR ILOCA	NON GENETIC	Post-radic encephalopathy
STARAC219	RARE GENETIC DISEASE OR ILOCA	NON GENETIC	Vascular
STARAC224	NON GENETIC	RARE GENETIC DISEASE OR ILOCA	ILOCA
STARAC229	NON GENETIC	RARE GENETIC DISEASE OR ILOCA	ILOCA
STARAC236	RARE GENETIC DISEASE OR ILOCA	NON GENETIC	Functional neurodisorders
STARAC237	ATX-FGF14/SCA27b	NON GENETIC	Post-radic encephalopathy
STARAC238	RARE GENETIC DISEASE OR ILOCA	NON GENETIC	Clinically probable MSA
STARAC243	ATX-FGF14/SCA27b	NON GENETIC	Autoimmune encephalitis

STARAC248	RARE GENETIC DISEASE OR ILOCA	NON GENETIC	Functional neurodisorders
STARAC256	RARE GENETIC DISEASE OR ILOCA	NON GENETIC	Clinically probable MSA
STARAC265	ATX-FGF14/SCA27b	RARE GENETIC DISEASE OR ILOCA	ILOCA
STARAC269	RARE GENETIC DISEASE OR ILOCA	NON GENETIC	Functional neurodisorders
STARAC271	ATX-FGF14/SCA27b	RARE GENETIC DISEASE OR ILOCA	ILOCA
STARAC279	RARE GENETIC DISEASE OR ILOCA	ATX-RFC1/CANVAS	ATX-RFC1/CANVAS
STARAC281	NON GENETIC	RARE GENETIC DISEASE OR ILOCA	ILOCA
STARAC285	RARE GENETIC DISEASE OR ILOCA	ATX-FGF14/SCA27b	ATX-FGF14b/SCA27b
STARAC292	RARE GENETIC DISEASE OR ILOCA	ATX-FGF14/SCA27b	ATX-FGF14b/SCA27b

2) External validation

ID	MODEL PREDICTION	ACCURATE PREDICTION	DIAGNOSIS DETAILS
STARAC10	RARE GENETIC DISEASE OR ILOCA	NON GENETIC	Functional neurodisorders
STARAC30	RARE GENETIC DISEASE OR ILOCA	NON GENETIC	Clinically probable MSA-C
STARAC52	NON GENETIC	RARE GENETIC DISEASE OR ILOCA	ILOCA
STARAC85	RARE GENETIC DISEASE OR ILOCA	NON GENETIC	Post-viral ataxia
STARAC99	NON GENETIC	RARE GENETIC DISEASE OR ILOCA	MELAS
STARAC109	NON GENETIC	RARE GENETIC DISEASE OR ILOCA	ILOCA
STARAC112	RARE GENETIC DISEASE OR ILOCA	ATX-FGF14/SCA27b	ATX-FGF14b/SCA27b
STARAC120	NON GENETIC	RARE GENETIC DISEASE OR ILOCA	ILOCA
STARAC132	RARE GENETIC DISEASE OR ILOCA	NON GENETIC	Functional neurodisorders
STARAC144	RARE GENETIC DISEASE OR ILOCA	ATX-FGF14/SCA27b	ATX-FGF14b/SCA27b
STARAC158	RARE GENETIC DISEASE OR ILOCA	NON GENETIC	Autoimmune encephalitis
STARAC159	ATX-FGF14/SCA27b	RARE GENETIC DISEASE OR ILOCA	ATX-ATN/DRPLA
STARAC167	RARE GENETIC DISEASE OR ILOCA	ATX-FGF14/SCA27b	ATX-FGF14b/SCA27b
STARAC181	RARE GENETIC DISEASE OR ILOCA	NON GENETIC	Chronic alcohol consumption
STARAC184	RARE GENETIC DISEASE OR ILOCA	NON GENETIC	Clinically probable MSA-C
STARAC199	RARE GENETIC DISEASE OR ILOCA	NON GENETIC	SMART syndrome
STARAC234	NON GENETIC	RARE GENETIC DISEASE OR ILOCA	ILOCA

Supplementary references

[1] K. Poirier *et al.*, « Mutations in TUBG1, DYNC1H1, KIF5C and KIF2A cause malformations of cortical development and microcephaly », *Nat. Genet.*, vol. 45, n° 6, p. 639-647, juin 2013, doi: 10.1038/ng.2613.

[2] J.-L. Méreaux *et al.*, « Fast and reliable detection of repeat expansions in spinocerebellar ataxia using exomes », *J. Med. Genet.*, vol. 60, n° 7, p. 717-721, juill. 2023, doi: 10.1136/jmg-2022-108924.