Non-reference genome transposable elements (TEs) have a significant impact on the progression of the Parkinson's disease

Journal:	*Experimental Biology and Medicine*
Manuscript ID	Draft
Manuscript Type:	Original Research
Date Submitted by the Author:	n/a
Complete List of Authors:	Koks, Sulev; Perron Institute for Neurological and Translational Science, ; Murdoch University, Bubb, Vivien; University of Liverpool, Singleton, Lewis; Perron Institute for Neurological and Translational Science Pfaff, Abigail; Murdoch University Quinn, John; University of Liverpool
Keywords:	Parkinson Disease, Clinical Study, Transposable Elements, Parkinson's Progression Markers Initiative, Longitudinal Study, Whole Genome Sequencing
Abstract:	The pathophysiology of Parkinson's disease (PD) is a complex process of the interaction between genetic and environmental factors. Studies on the genetic component of PD have predominantly focused on single nucleotide polymorphisms (SNP) using a cross-sectional case-control design in large genome-wide association studies. This approach whilst giving insight into a significant portion of the genetics of PD does not fully account for all the genetic components resulting in missing heritability. In the present study, we approached this problem by focusing on the non-reference genome transposable elements (TE) and their impact on the progression of PD using a longitudinal study design within the Parkinson's Progression Markers Initiative (PPMI) cohort. We analysed 2,886 Alu repeats, 360 LINE1 and 128 SVAs that were called from the whole genome sequence data which are not within the reference genome. The presence or absence of these non-reference TE variants is known as a retrotransposon insertion polymorphism and measuring this polymorphism describes the impact of TEs on the traits. The variations for the presence or absence of the non-reference TE elements were modelled to align with the changes in the 114 outcome measures during the five-year follow-up period of the PPMI cohort. Linear mixed-effects models were used and many TEs were found to have a highly significant effect on the longitudinal changes in the clinically important PD outcomes such as UPDRS subscale II, UPDRS total scores and modified Schwab and England ADL scale. In addition, the progression of several imaging and functional measures, including the Caudate/Putamen ratio and levodopa equivalent daily dose (LEDD) were also significantly affected by the TEs. In conclusion, this study identified the overwhelming effect of the non-reference TEs on the progression of PD and is a good example of the impact the variations in the "junk DNA" have on complex diseases.

Non-reference genome transposable elements (TEs) have a significant impact on the progression of the Parkinson's disease

Sulev Kõks [1,2,*], Abigail L. Pfaff [1,2], Lewis M. Singleton [1], Vivien J. Bubb [3] and John P. Quinn [3]

[1] Perron Institute for Neurological and Translational Science, Perth, WA 6009, Australia

[2] Centre for Molecular Medicine and Innovative Therapeutics, Murdoch University, Perth, WA 6150, Australia

[3] Department of Pharmacology and Therapeutics, Institute of Systems, Molecular and Integrative Biology, University of Liverpool, Liverpool L69 3BX, UK

[*] Correspondence: sulev.koks@perron.uwa.edu.au; Tel.: +61-(0)-8-6457-0313

Abstract

The pathophysiology of Parkinson's disease (PD) is a complex process of the interaction between genetic and environmental factors. Studies on the genetic component of PD have predominantly focused on single nucleotide polymorphisms (SNP) using a cross-sectional case-control design in large genome-wide association studies. This approach whilst giving insight into a significant portion of the genetics of PD does not fully account for all the genetic components resulting in missing heritability. In the present study, we approached this problem by focusing on the non-reference genome transposable elements (TE) and their impact on the progression of PD using a longitudinal study design within the Parkinson's Progression Markers Initiative (PPMI) cohort. We analysed 2,886 Alu repeats, 360 LINE1 and 128 SVAs that were called from the whole genome sequence data which are not within the reference genome. The presence or absence of these non-reference TE variants is known as a retrotransposon insertion polymorphism and measuring this polymorphism describes the impact of TEs on the traits. The variations for the presence or absence of the non-reference TE elements were modelled to align with the changes in the 114 outcome measures during the five-year follow-up period of the PPMI cohort. Linear mixed-effects models were used and many TEs were found to have a highly significant effect on the longitudinal changes in the clinically important PD outcomes such as UPDRS subscale II, UPDRS total scores and modified Schwab and England ADL scale. In addition, the progression of several imaging and functional measures, including the Caudate/Putamen ratio and levodopa equivalent daily dose (LEDD) were also significantly affected by the TEs. In conclusion, this study identified the overwhelming effect of the non-reference TEs on the

progression of PD and is a good example of the impact the variations in the “junk DNA” have on complex diseases.

Keywords

Parkinson Disease, Clinical Study, Transposable Elements, Parkinson's Progression Markers Initiative, Longitudinal Study, Whole Genome Sequencing,

Impact Statement

The present study analysed the genomic variation in the "dark matter" or noncoding component of the genome and its impact on the progression of Parkinson's disease (PD). We demonstrate the presence or absence of non-reference transposable elements (TEs) in the human genome modifies significantly the longitudinal clinical course of the PD and the progression of the neurodegeneration in the brain of the patients. The effect of TE can be either protective or damaging for the PD progression. This finding has a significant impact on our understanding of the role of the TEs in PD and presents the need to redefine the function of genomic repetitive elements that form the largest component of the human genome.

Introduction

More than 70% of human genome consists of repetitive transposable elements (TEs) much of which does not encode any proteins and was therefore generally considered as a useless parasitic vestige of the past viral infections[1]. However, many recent studies indicate that this part of the genome has quite a significant impact on genome regulation and function and requires more rigorous and targeted analysis[2-6]. These functional studies have shown that TEs like Alu repeats, LINE1s and composite SINE-VNTR-Alu (SVA) can modify gene expression and be a part of the disease mechanism [7, 8]. Most importantly, analysis of the variability of the TEs offers a unique opportunity to identify the hidden genetic mechanisms of disease, often referred simply as the missing heritability [9]. Genome wide association studies (GWAS) based on SNP genotyping have enjoyed significant success in defining the genetics of complex diseases. Nevertheless, these findings have repeatedly been reported to suffer from severe limitations starting with the very low effect sizes (odds ratios usually below 1.5), lack of predictive power of SNPs identified and the limits of explaining the heritability of diseases even by the very large meta-analyses[10]. GWAS detected odds ratios below 1.5 can mostly be explained by cryptic population stratification regardless of the p-value[11]. The limits of GWAS studies have led to the missing heritability problem, the gap between the amount of heritability that can be explained by GWAS and the amount that is estimated from twin studies [12]. This is where the analysis of structural variation and TEs would potentially supply an additional layer of genomic variation that might in part explain this hidden variability of the human genome.

Previous studies have already shown the feasibility of the analysis of TEs in the context of complex diseases [13-17]. The most common type of the polymorphism to be studied in TEs is their presence or absence in the genome. X-linked dystonia parkinsonism

has been found to be caused by a SVA insertion in the intron of the TAF1 gene that modulated expression of the TAF1 and D2 receptor gene in the caudate nucleus [15]. In addition, removal of the SVA insertion from the intron restored TAF1 expression to the normal levels indicating the causal link between the SVA and expression regulation. Moreover, this finding illustrates potential therapeutic approach of excising the SVA element from the genome or modifying the activity of SVA [18]. Bardet-Biedl syndrome is another example illustrating the role of TEs in the pathogenesis of disease. A SVA F insertion was recently identified in the exon 13 of the BBS1 gene as a causative mutation for several families with the syndrome [19, 20]. Our own recent study showed the involvement of polymorphic reference genome SVAs in the neurodegeneration and the progression of Parkinson's disease [13]. Moreover, we and other groups have also shown that TEs have significant, large, and genome-wide effect on gene expression [21, 22]. The regulatory effect of TEs on *cis* or *trans* gene expression helps to explain the mechanism of the elements on the disease risk and progression [13, 14]. At the same time, presence of the TEs in the genome can induce alternative splicing, exonisation, intron retention, transcript fusion or premature stop that all can lead to the disease [8]. This makes TEs targets to identify new genomic loci or genetic elements responsible for the development of diseases and a new class of therapeutic targets.

However, our previous study focused on the retrotransposon insertion polymorphisms (RIP) described in the reference genome. We decided to have a different approach and use non-reference genome TE polymorphisms as markers to analyse the genomic variants responsible for the Parkinson's disease (PD). The Longitudinal Parkinson's Progression Markers Initiative (PPMI) cohort offers a unique opportunity for this type of studies as it combines rich clinical information, drug response, imaging and biochemical

data taken repeatedly from the same patients over at least five years. At the same time, whole genome sequencing data with an annual blood transcriptome snapshot is available for every individual. This design leverages the real-world data by incorporating heterogenous population of PD patients with their natural course of the disease. Therefore, PPMI dataset is the most suitable to analyse the impact of non-reference TEs on the course of PD.

Materials and Methods

Study cohort

Only PD patients' data of the Parkinson's Progression Markers Initiative (PPMI) was used in the longitudinal analysis. Data of control subjects were not used as the goal of the study was to analyse the effect of TEs of the progression of PD. Briefly, PPMI contains longitudinal data of 423 PD patients with 157 different clinical, imaging, or biochemical traits, that all were used for our initial analysis. After quality control and removing the non-variable and uninformative data, 114 traits were eventually used for longitudinal analysis.

Identifying non-reference TE presence/absence polymorphisms and association analysis

Whole genome sequencing (WGS) data were obtained from PPMI in BAM file format. Mobile element locator tool (MELT version 2.1.5 in MELT-split mode) was used to call and genotype non-reference *Alu*, L1 and SVA non-LTR retrotransposons for their presence/absence in 1336 genomes using Pawsey supercomputing infrastructure [23]. This included 191 healthy controls, 394 PD subjects, 63 individuals with scans without evidence of dopaminergic deficit (SWEDD), 63 prodromal individuals and 625 individuals harbouring a known genetic variant associated with PD (360 unaffected individuals and 265 with PD). The retrotransposon variants detected were filtered to keep those supported by >2 split reads and assess score ≥3 and that had passed the filtering criteria performed by MELT. Disease association analysis was performed on 3375 retrotransposons variants (2887 *Alu*, 360 L1 and 128 SVAs) that were in Hardy-Weinberg equilibrium (variants removed if $p<1x10^{-6}$ in healthy controls) and had an insertion allele frequency greater than 0.01 in the healthy controls and PD subjects. Logistic regression with sex, age, ethnicity, and family history as covariates was performed using genotypes from the healthy control and PD subjects in Plink (v1.07). The p-values were corrected for multiple tested using Bonferroni [24].

The use of longitudinal PPMI clinical and genomic data was approved by the Human ethics committee of the Murdoch University.

Analysis of the TE effects on the progression of PD

Linear mixed effects modelling was used to analyse the effect of presence or absence polymorphism of TEs on the clinical traits. The modelling was performed in the R studio and with the *LmerTest* R package. Following formula for longitudinal modelling of the effect of the non-reference TEs on the change of the trait between visits was used:

anova(lmerTest::lmer(TRAIT ~ NON-REF-TE * months + (1|PATNO),

na.action=na.omit,data=PD))

Resulting P-values were FDR adjusted for the multiple correction and only FDR values below 0.05 were considered statistically significant. Corrected FDR values were used to select significant traits for the pairwise analysis of the effects of each genotype using *emmeans* package. For Manhattan plotting FDR corrected p-values were used and the plots show FDR corrected values.

Results

Association analysis

In 1,336 whole genomes from the PPMI cohort 16,438 non-reference retrotransposon insertions were detected, consisting of 13,041 Alus, 2354 L1s and 1,043 SVAs. Most of the insertions detected were rare and more than half had an insertion allele frequency (IAF) of <0.001 (Figure 1a). Insertions were predominantly located in either intergenic or intronic regions and small number were in exons, untranslated regions, and promoters (Figure 1b). SVAs were more frequently located in introns and promoters compared to Alus and L1s. Association analysis was performed on those insertions (3,375) with an IAF greater than 0.01 in the healthy controls (191 individuals) and PD subjects (394 individuals). After correction for multiple testing there were no retrotransposon insertions associated with an increased risk of developing PD.

Longitudinal analysis

We analysed the effect of the 3,374 non reference TEs on the PD progression in the 423 patients with data for five visits, baseline followed by four annual follow ups, on 114 traits. The numbers of different TE types used for longitudinal analysis are illustrated in the Figure 2. Out of 3,374 TEs, 1,581 Alu repeats, 205 LINE1 elements and 75 SVA elements gave significant effect on at least one clinical, biochemical, or imaging trait measured in the PPMI cohort.

Different TEs had different effects on PD phenotypes and progression. From all non-reference 2,886 Alu elements, 1,305 were without any effect, from all 360 LINE1 elements, 155 were without longitudinal effect and from all detected 128 SVAs, 53 were without any effect. The elements with most frequent association with the PD progression traits were NR-Alu-1388 (18 traits), NR-L1-1126 (18 traits) and NR-SVA-982 (10 traits). At the same time,

different traits were associated with the variable numbers of TEs with the statistically significant effects (Figure 3). For all TEs, the most commonly affected traits were primary diagnosis (268 hits), UPDRS Part II score (224 hits), UDPRS Total Score ON (213), Modified Schwab & England ADL Score (MSEADLG, 190 hits) UDPRS Total Score OFF (161), left side DaTscan Caudate/Putamen Count Density Ratio (l-CDR, 150 hits), Sexual Impulse Control Disorder (QUIP-sex, 145 hits), change in diagnosis (138 hits) LEDD (120 hits) and ipsilateral CDR (ips-CDR, 119 hits). Different TEs had slightly different profiles in the trait they modulate. Three most affected traits for Alu repeats were primary diagnosis, UPDRS Part II score and UDPRS Total Score ON. Primary diagnosis is the primary diagnosis at the time of recruitment, and it could change during the follow-ups. The most common traits affected by the LINE1 elements were UPDRS Part II score, primary diagnosis, and ips-CDR (ipsilateral count density ratio of Caudate/Putamen). SVAs preferably modified Symbol Digit Modalities Score (SDMTOTAL, test for cognitive impairment), primary diagnosis, the change in primary diagnosis and Modified Schwab and England ADL score (MSEADLG). Manhattan plots in the following figures indicate the location and p-values of the TEs affecting changes in the traits during the follow-up period. The Figure 4 shows all TEs affecting the SDMTOTAL, Figure 5 shows the elements affecting UPDRS Part II score with their genomic location and Figure 6 shows FDR values and positions of all the TEs affecting Levodopa Equivalent Daily Dose (LEDD).

We next analysed the specific effects that the specific TEs had on the progression traits. Figure 7 illustrates the change in UPDRS Part II score between different visit and its dependency on the NR-Alu-10169 genotype. Interestingly, patients with the absence (AA) of NR-Alu-10169 progressed significantly faster at visits 8, 10 and 12compared to the same visits of different genotypes (PA and PP). This shows the protective effect of that specific Alu

element on the progression of PD. Figure 8 illustrates another Alu element, NR-Alu-8491 and its effect on UPDRS Total ON score. Compared to AA and PA, patients with PP genotype had significantly higher scores at visits 8, 10 and 12. Importantly, the difference between different genotypes is almost 40 points that indicates very large clinically important difference related to the presence of the NR-Alu-8491 repeat [25].

Figures 9 and 10 illustrate some effects of the LINE1 elements on the PD progression. Figure 9 shows faster progression of UPDRS Part II score in patients with PP genotype for NR-L1-1126 and again the significant difference emerged from visit 8 onwards. Differences in UPDRS II scores were more than 20 points indicating again very large clinical significance in patients with PP NR-L1-1126 genotypes. Figure 10 illustrates that the degeneration of the putamen depends on the NR-L1-1652 genotype and presents decreased putamen volume in patients without this element (AA). Putamen volume was measured as the Caudate/Putamen count density ratio (l-CDR) and increase in this ratio shows degeneration of the Putamen. Interestingly, presence of the NR-L1-1652 element was protective as no change over five years was detected in patients with PA or PP genotypes, showing even single copy of NR-L1-1652 being protective against the putamen degeneration.

Finally, to describe the effect of SVA, we show its role in the cognitive decline of PD (Supplementary Figure 1). The presence of NR-SVA-365 PP genotype, was significantly related to the accelerated cognitive decline in PD patients as measured with the UPDRS Part I Cognitive Impairment score. Figure 11 shows clear increase in the NP1COG scores of the patients and this change is significant from the visit 11 onwards.

Taken together, we identified many non-reference TEs to have an impact on the progression of PD, most remarkably on the progression of the UPDRS subscores and degeneration of the putamen.

Discussion

We have analysed the presence/absence polymorphisms of the non-reference TEs and the progression of PD in a longitudinal study involving of 423 patients followed up for five years. We focused on the Alu, LINE1 and SVA elements the non-LTR retrotransposon component representing ~35% of the human genome subgroup of TEs. We identified highly significant genetic influence of these elements on the PD progression, and this was evident in many different clinical, imaging, and biochemical traits.

We identified that TEs correlate significantly with the progression of several clinically important traits of PD patients including LEDD, UPDRS total and sub scores and cognition. These genomic associations were not clustered to one location but were rather spread out into different distinct genomic locations. Figures 4, 5, and 6 illustrate with Manhattan plots the spread and significance of the elements on the specific PD traits. These figures also show different TEs influencing the same trait can be very close to each other in the genome suggesting potential functional underlying locus the TE is regulating. One mechanism this genomic co-localisation implies is the quantitative trait locus or QTL, that is the region in the genome involved in the development of the phenotypic outcome. TEs have been described to have genome wide eQTL effect and they are known to have very large quantitative effect size in the gene expression [21, 22, 26]. Changes in the transcriptome can involve both transcriptional and post transcriptional mechanism including for the alternative splicing or intronic retentions. In our recent study, we analysed the intronic transcripts and demonstrated widespread nascent transcription in the context of PD [27]. Therefore, it is quite likely that TEs form regulatory sites that dictate the splicing or other transcriptional changes that in part are relevant for the disease as shown previously [28]. The hypothesis that TEs are correlated with PD progression needs further testing with the studies involving tissue

samples from the patients. While the primary pathology of PD is in the brain, to leverage the power of the longitudinal design, blood transcriptome would provide excellent information about the potential pathological changes. We have shown the viability of using peripheral tissue as a surrogate tissue for brain disease before [29]. Prospective cohort design is the gold standard design for clinical research to detect the causative relations between the clinical and genetic variables. Therefore, the PPMI cohort we have used in the present study, offers invaluable opportunity to identify the missing heritability component for the PD.

The most unexpected finding is that presence or absence of TEs influences the primary diagnosis of the disease that we measured as a primary diagnosis and as a change of primary diagnosis. Patients with certain TEs were more likely to receive incorrect diagnosis at the early visits that was corrected later. The change of diagnosis could indicate the difference in the endophenotypes of the subjects with different TE genotypes. It is important to stress here, that accurate diagnosis at the beginning of the PD is challenging because of the complexity of the phenotype and overlapping extrapyramidal syndromes [30, 31]. The finding that some of the TEs are associated with the significantly higher frequency of diagnostic challenges, might reflect genetic heterogeneity of the PD and could suggest even a separate subtype of the disease.

Imaging of the brain structures is possibly one of the best and the most reliable measure to characterise PD and its progression. As with the initial diagnosis of PD, when the symptoms of the disease can be very different between patients, the progression of PD and concomitant degeneration in the imaging is also individually highly variable. Many TEs had very clear association with the CDR measure that is calculated as the Caudate/Putamen ratio using DatScan data in the PPMI cohort. Increase in this ratio shows decrease in the Putamen volume, and in several cases, we identified changes in CDR to be related to the

specific different genotypes of TEs. Dopaminergic degeneration in PD is not uniform regionally and the Putamen is affected more than other regions [32]. This feature is helpful to differentiate idiopathic PD from atypical PD forms [33]. Putamen dopaminergic dysfunction has been shown to be the best predictive risk factor for the REM sleep behaviour phenoconversion to the overt synucleinopathy [34]. Moreover, a greater reduction in the Putamen dopaminergic binding in relation to the caudate has been shown to be specific for the PD compared to the traumatic brain injury [35]. Therefore, the increase in the Caudate/Putamen ratio is an indication of the progression of the dopaminergic neurodegeneration in the brains of PD patients. In the present study, genetic variations in at least 150 TE elements were found to be related to the faster degeneration of the Putamen and faster progression of PD. This is a strong indication that polymorphic TE loci are directly related to the neuropathology changes during the progression of the PD.

In conclusion, the present study described the significant impact of non-reference TEs in the progression of PD and in its specific traits. Our main finding is that the presence or absence of TEs changes progression trajectory of PD and we provided clinical, imaging, and biochemical evidence to support this. Non-reference TEs are involved in the regionally specific dopaminergic neurodegeneration of the putamen while preserving other parts of striatum and that might be the leading cause connecting changes in the other traits described in this study. Our study will not have captured the complete repertoire of non-reference genome TEs due to the difficulty in characterising these elements in short read sequence data, indicating the wealth of genetic information associated with disease to be determined from a rigorous analysis of these elements in our genome.

Authors' Contributions

Conceptualization, S.K.; methodology, A.L.P., L.M.S. and S.K.; formal analysis, S.K.; data interpretation, A.L.P., L.M.S., V.J.B., J.P.Q. and S.K.; writing—original draft preparation, S.K.; writing—review and editing, A.L.P., L.M.S., V.J.B., J.P.Q. and S.K.; funding acquisition, A.L.P. and S.K. All authors have read and agreed to the published version of the manuscript.

Acknowledgements

This work was supported by resources provided by the Pawsey Supercomputing Centre with funding from the Australian Government and the Government of Western Australia. Data used in the preparation of this article were obtained from the Parkinson's Progression Markers Initiative (PPMI) database (www.ppmi-info.org/data (accessed on 19 January 2021)). For up-to-date information on the study, visit www.ppmi-info.org. PPMI is sponsored and partially funded by The Michael J. Fox Foundation for Parkinson's Research. PPMI—a public-private partnership—Is funded by the Michael J. Fox Foundation for Parkinson's Research and funding partners, including Abbvie, Allergan, Amathus Therapeutics, Avid Radiopharmaceuticals, Biogen Idec, Biolegend, Briston-Myers Squibb, Celgene, Denali, GE Healthcare, Genentech, GlaxoSmithKline, Janssen neuroscience, Lilly, Lundbeck, Merck, Meso Scale Discovery, Pfizer, Piramal, Prevail Therapeutics, Roche, Sanofi Genzyme, Servier, Takeda, Teva, UCB, Verily and Voyager Therapeutics.

Declaration of Conflicting Interests

The authors declare no conflict of interest related to this study. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Ethical Approval

The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Institutional Human Ethics Research Office of The University of Western Australia (protocol code RA/4/20/5308 approved on 05.08.2019)

Funding

This research was funded by MSWA, The Michael J. Fox Foundation, Shake It Up Australia and Perron Institute for Neurological and Translational Science.

Figure Legends and References.

Figure 1. Allele frequency and location of non-reference retrotransposon insertion polymorphisms in the PPMI cohort. A) The percentage of non-reference retrotransposon insertion polymorphisms with certain allele frequencies in the PPMI cohort. B) The percentage non-reference retrotransposon polymorphisms located in specific regions of the genome.

Figure 2. Overview of all the non-reference TEs we used in the longitudinal analysis.

Figure 3. Number of the statistically significant non-reference TEs (Alu, L1 or SVA) related to the changes in the traits during the progression of PD. The larger number indicates higher number of the TEs modifying the respective trait.

Figure 4. Manhattan plot of all non-ref TEs with their location in the genome and their FDR values for the longitudinal changes of the Symbol Digit Modalities Score (SDMTOTAL). Only the most significant (FDR below 0.001) are labelled, Y-axis value is -log10 of the FDR corrected p-value.

Figure 5. Manhattan plot of all non-ref TEs with their location in the genome and the FDR values for the longitudinal changes of the UPDRS Part 2 score (UPDRS2). Only the most significant (FDR below 0.001) are labelled, Y-axis value is -log10 of the FDR corrected p-value.

Figure 6. Manhattan plot of all non-ref TEs with their location in the genome and the FDR values for the longitudinal changes of the Total Levodopa Equivalent Daily Dose (LEDD). Only the most significant (FDR below 0.001) are labelled, Y-axis value is -log10 of the FDR corrected p-value.

Figure 7. Longitudinal changes in the UPDRS Part 2 scores and the differences related to the presence or absence of the non-ref Alu-10169. Corrected p-values are presented as * - p-value < 0.05, *** - p-value < 0.001.

Figure 8. Longitudinal changes in the UPDRS Total Scores and the differences related to the presence or absence of the non-ref Alu-8491. Corrected p-values are presented as *** - p-value < 0.001.

Figure 9. Longitudinal changes in the UPDRS Part 2 scores and the differences related to the presence or absence of the non-ref LINE1-1126. Corrected p-values are presented as *** - p-value < 0.001.

Figure 10. Longitudinal changes in the Caudate/Putamen Ratio (IPS-CDR) and the differences in these changes related to the presence or absence of the non-ref LINE1-1652. Increased ratio indicates progressive degeneration of the Putamen from V04 and during the consecutive visits. Corrected p-values are presented as *** - p-value < 0.001.

Figure 11. Longitudinal progression of the cognitive impairment measured by the UPDRS Cognitive scores and the differences in the progression related to the presence or absence of the non-ref SVA-365.Corrected p-values are presented as *** - p-value < 0.001.

Supplementary Figure 1. Manhattan plot of all non-ref TEs with their location in the genome and the FDR values for the longitudinal changes of the cognitive impairment measured with the UPDRS Part I (NP1COG). Only the most significant (FDR below 0.001) are labelled, Y-axis value is -log10 of the FDR corrected p-value.

References

1. de Koning AP, Gu W, Castoe TA, Batzer MA, Pollock DD. Repetitive elements may comprise over two-thirds of the human genome. *PLoS Genet* 2011;**7**:e1002384
2. Ridker PM, Baker MT, Hennekens CH, Stampfer MJ, Vaughan DE. Alu-repeat polymorphism in the gene coding for tissue-type plasminogen activator (t-PA) and risks of myocardial infarction among middle-aged men. *Arteriosclerosis, thrombosis, and vascular biology* 1997;**17**:1687-90
3. Belancio VP, Deininger PL, Roy-Engel AM. LINE dancing in the human genome: transposable elements and disease. *Genome medicine* 2009;**1**:97
4. Alexander RP, Fang G, Rozowsky J, Snyder M, Gerstein MB. Annotating non-coding regions of the genome. *Nature reviews Genetics* 2010;**11**:559-71
5. Akrami SM, Habibi L. Retrotransposons and pediatric genetic disorders: Importance and implications. *Journal of pediatric genetics* 2014;**3**:9-16
6. Palazzo AF, Gregory TR. The case for junk DNA. *PLoS Genet* 2014;**10**:e1004351
7. Marshall JN, Lopez AI, Pfaff AL, Koks S, Quinn JP, Bubb VJ. Variable number tandem repeats - Their emerging role in sickness and health. *Exp Biol Med (Maywood)* 2021;**246**:1368-76
8. Pfaff AL, Singleton LM, Koks S. Mechanisms of disease-associated SINE-VNTR-Alus. *Exp Biol Med (Maywood)* 2022:15353702221082612
9. Hancks DC, Kazazian HH, Jr. Active human retrotransposons: variation and disease. *Curr Opin Genet Dev* 2012;**22**:191-203
10. Tam V, Patel N, Turcotte M, Bosse Y, Pare G, Meyre D. Benefits and limitations of genome-wide association studies. *Nature reviews Genetics* 2019;**20**:467-84
11. McClellan J, King MC. Genetic heterogeneity in human disease. *Cell* 2010;**141**:210-7
12. Keller MF, Saad M, Bras J, Bettella F, Nicolaou N, Simón-Sánchez J, Mittag F, Büchel F, Sharma M, Gibbs JR, Schulte C, Moskvina V, Durr A, Holmans P, Kilarski LL, Guerreiro R, Hernandez DG, Brice A, Ylikotila P, Stefánsson H, Majamaa K, Morris HR, Williams N, Gasser T, Heutink P, Wood NW, Hardy J, Martinez M, Singleton AB, Nalls MA. Using genome-wide complex trait analysis to quantify 'missing heritability' in Parkinson's disease. *Hum Mol Genet* 2012;**21**:4996-5009
13. Pfaff AL, Bubb VJ, Quinn JP, Koks S. Reference SVA insertion polymorphisms are associated with Parkinson's Disease progression and differential gene expression. *NPJ Parkinsons Dis* 2021;**7**:44
14. Pfaff AL, Bubb VJ, Quinn JP, Koks S. An Increased Burden of Highly Active Retrotransposition Competent L1s Is Associated with Parkinson's Disease Risk and Progression in the PPMI Cohort. *Int J Mol Sci* 2020;**21**
15. Makino S, Kaji R, Ando S, Tomizawa M, Yasuno K, Goto S, Matsumoto S, Tabuena MD, Maranon E, Dantes M, Lee LV, Ogasawara K, Tooyama I, Akatsu H, Nishimura M, Tamiya G. Reduced neuron-specific expression of the TAF1 gene is associated with X-linked dystonia-parkinsonism. *American journal of human genetics* 2007;**80**:393-406
16. Aneichyk T, Hendriks WT, Yadav R, Shin D, Gao D, Vaine CA, Collins RL, Domingo A, Currall B, Stortchevoi A, Multhaupt-Buell T, Penney EB, Cruz L, Dhakal J, Brand H, Hanscom C, Antolik C, Dy M, Ragavendran A, Underwood J, Cantsilieris S, Munson KM, Eichler EE, Acuna P, Go C, Jamora RDG, Rosales RL, Church DM, Williams SR, Garcia S, Klein C, Muller U, Wilhelmsen KC, Timmers HTM, Sapir Y, Wainger BJ, Henderson D, Ito N, Weisenfeld N, Jaffe

D, Sharma N, Breakefield XO, Ozelius LJ, Bragg DC, Talkowski ME. Dissecting the Causal Mechanism of X-Linked Dystonia-Parkinsonism by Integrating Genome and Transcriptome Assembly. *Cell* 2018;**172**:897-909 e21

17. Bragg DC, Mangkalaphiban K, Vaine CA, Kulkarni NJ, Shin D, Yadav R, Dhakal J, Ton ML, Cheng A, Russo CT, Ang M, Acuna P, Go C, Franceour TN, Multhaupt-Buell T, Ito N, Muller U, Hendriks WT, Breakefield XO, Sharma N, Ozelius LJ. Disease onset in X-linked dystonia-parkinsonism correlates with expansion of a hexameric repeat within an SVA retrotransposon in TAF1. *Proceedings of the National Academy of Sciences of the United States of America* 2017;**114**:E11020-E28

18. Rakovic A, Domingo A, Grutz K, Kulikovskaja L, Capetian P, Cowley SA, Lenz I, Bruggemann N, Rosales R, Jamora D, Rolfs A, Seibler P, Westenberger A, Konig I, Klein C. Genome editing in induced pluripotent stem cells rescues TAF1 levels in X-linked dystonia-parkinsonism. *Mov Disord* 2018;**33**:1108-18

19. Tavares E, Tang CY, Vig A, Li S, Billingsley G, Sung W, Vincent A, Thiruvahindrapuram B, Heon E. Retrotransposon insertion as a novel mutational event in Bardet-Biedl syndrome. *Mol Genet Genomic Med* 2019;**7**:e00521

20. Delvallee C, Nicaise S, Antin M, Leuvrey AS, Nourisson E, Leitch CC, Kellaris G, Stoetzel C, Geoffroy V, Scheidecker S, Keren B, Depienne C, Klar J, Dahl N, Deleuze JF, Genin E, Redon R, Demurger F, Devriendt K, Mathieu-Dramard M, Poitou-Bernert C, Odent S, Katsanis N, Mandel JL, Davis EE, Dollfus H, Muller J. A BBS1 SVA F retrotransposon insertion is a frequent cause of Bardet-Biedl syndrome. *Clin Genet* 2021;**99**:318-24

21. Koks S, Pfaff AL, Bubb VJ, Quinn JP. Expression Quantitative Trait Loci (eQTLs) Associated with Retrotransposons Demonstrate their Modulatory Effect on the Transcriptome. *Int J Mol Sci* 2021;**22**

22. Wang L, Rishishwar L, Marino-Ramirez L, Jordan IK. Human population-specific gene expression and transcriptional network modification with polymorphic transposable elements. *Nucleic acids research* 2017;**45**:2318-28

23. Gardner EJ, Lam VK, Harris DN, Chuang NT, Scott EC, Pittard WS, Mills RE, Devine SE. The Mobile Element Locator Tool (MELT): population-scale mobile element discovery and biology. *Genome research* 2017;**27**:1916-29

24. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. *American journal of human genetics* 2007;**81**:559-75

25. Shulman LM, Gruber-Baldini AL, Anderson KE, Fishman PS, Reich SG, Weiner WJ. The clinically important difference on the unified Parkinson's disease rating scale. *Arch Neurol* 2010;**67**:64-70

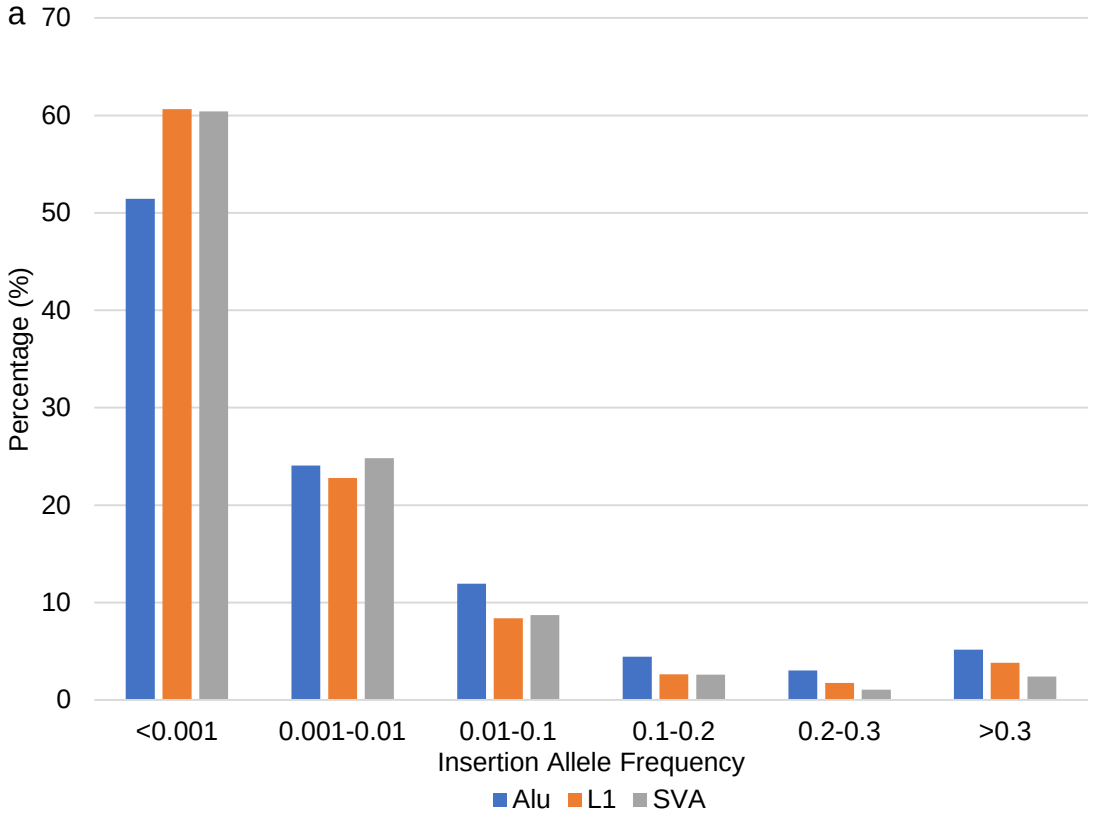
26. Wang Q, Zhang Y, Wang M, Song WM, Shen Q, McKenzie A, Choi I, Zhou X, Pan PY, Yue Z, Zhang B. The landscape of multiscale transcriptomic networks and key regulators in Parkinson's disease. *Nature communications* 2019;**10**:5234

27. Koks S, Pfaff AL, Bubb VJ, Quinn JP. Longitudinal intronic RNA-Seq analysis of Parkinson's disease patients reveals disease-specific nascent transcription. *Exp Biol Med (Maywood)* 2022;**247**:945-57

28. Guelfi S, D'Sa K, Botía JA, Vandrovcova J, Reynolds RH, Zhang D, Trabzuni D, Collado-Torres L, Thomason A, Quijada Leyton P, Gagliano Taliun SA, Nalls MA, Small KS, Smith C, Ramasamy A, Hardy J, Weale ME, Ryten M. Regulatory sites for splicing in human basal ganglia are enriched for disease-relevant information. *Nature communications* 2020;**11**:1041

29. Lill M, Koks S, Soomets U, Schalkwyk LC, Fernandes C, Lutsar I, Taba P. Peripheral blood RNA gene expression profiling in patients with bacterial meningitis. *Front Neurosci* 2013;**7**:33
30. Zhang PL, Chen Y, Zhang CH, Wang YX, Fernandez-Funez P. Genetics of Parkinson's disease and related disorders. *J Med Genet* 2018;**55**:73-80
31. Halli-Tierney AD, Luker J, Carroll DG. Parkinson Disease. *Am Fam Physician* 2020;**102**:679-91
32. Jaimini A, Tripathi M, D'Souza MM, Panwar P, Sharma R, Mehta S, Pandey S, Saw S, Singh D, Solanki Y, Mishra AK, Mondal A. Utility of intrastriatal ratios of FDOPA to differentiate idiopathic Parkinson's disease from atypical parkinsonian disorders. *Nucl Med Commun* 2013;**34**:426-31
33. Stormezand GN, Chaves LT, Vállez García D, Doorduin J, De Jong BM, Leenders KL, Kremer BPH, Dierckx R. Intrastriatal gradient analyses of 18F-FDOPA PET scans for differentiation of Parkinsonian disorders. *Neuroimage Clin* 2020;**25**:102161
34. Arnaldi D, Chincarini A, Hu MT, Sonka K, Boeve B, Miyamoto T, Puligheddu M, De Cock VC, Terzaghi M, Plazzi G, Tachibana N, Morbelli S, Rolinski M, Dusek P, Lowe V, Miyamoto M, Figorilli M, Verbizier D, Bossert I, Antelmi E, Meli R, Barber TR, Trnka J, Miyagawa T, Serra A, Pizza F, Bauckneht M, Bradley KM, Zogala D, McGowan DR, Jordan L, Manni R, Nobili F. Dopaminergic imaging and clinical predictors for phenoconversion of REM sleep behaviour disorder. *Brain* 2021;**144**:278-87
35. Jenkins PO, Roussakis AA, De Simoni S, Bourke N, Fleminger J, Cole J, Piccini P, Sharp D. Distinct dopaminergic abnormalities in traumatic brain injury and Parkinson's disease. *Journal of neurology, neurosurgery, and psychiatry* 2020;**91**:631-37

Figure 1


a
70
60
50
40
30
20
10
0
Percentage (%)
<0.001
0.001-0.01
0.01-0.1
0.1-0.2
0.2-0.3
>0.3
Insertion Allele Frequency
Alu
L1
SVA

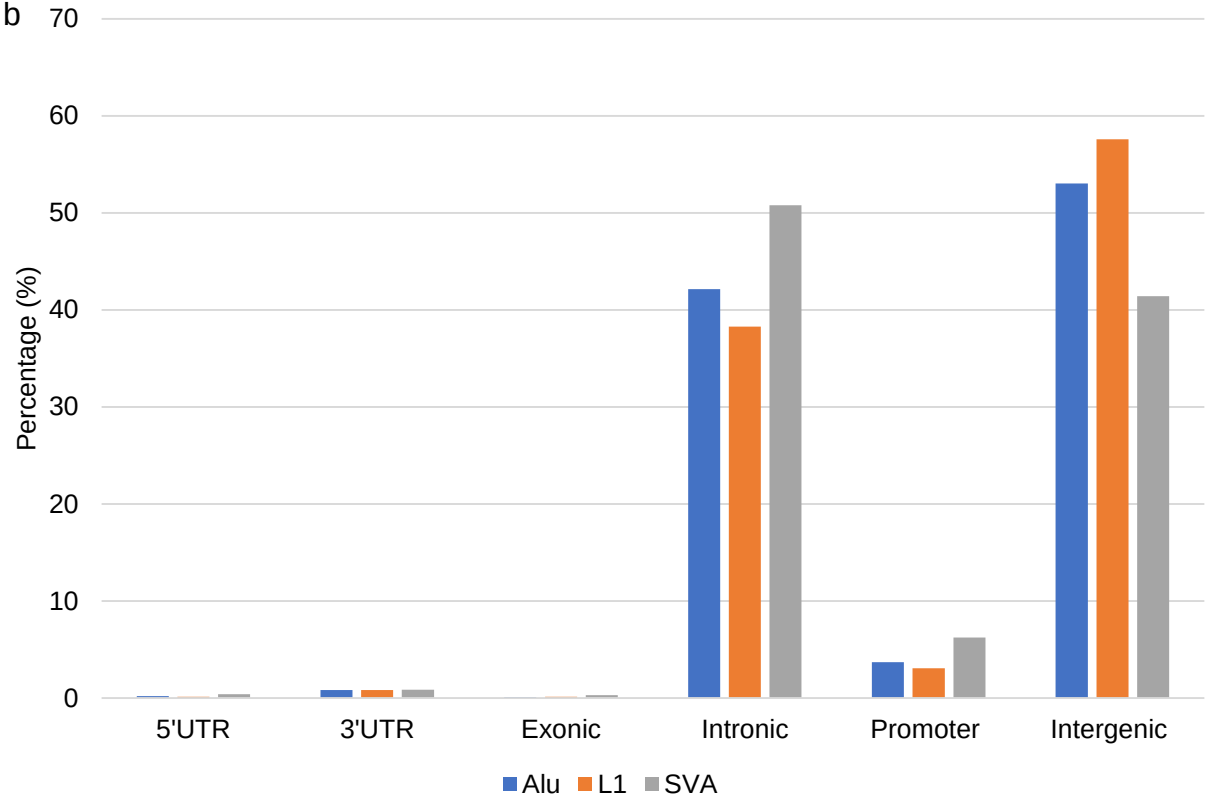
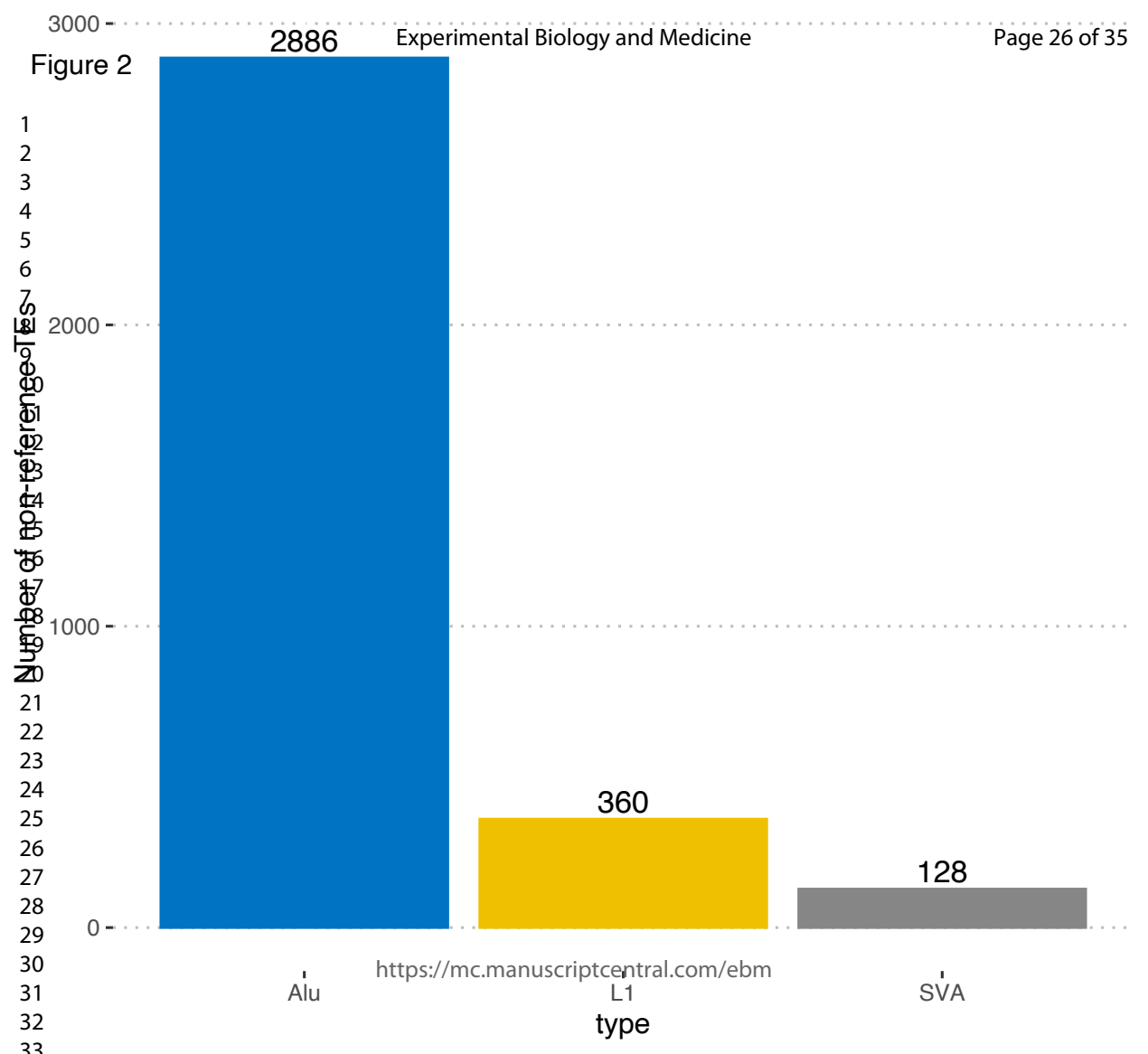
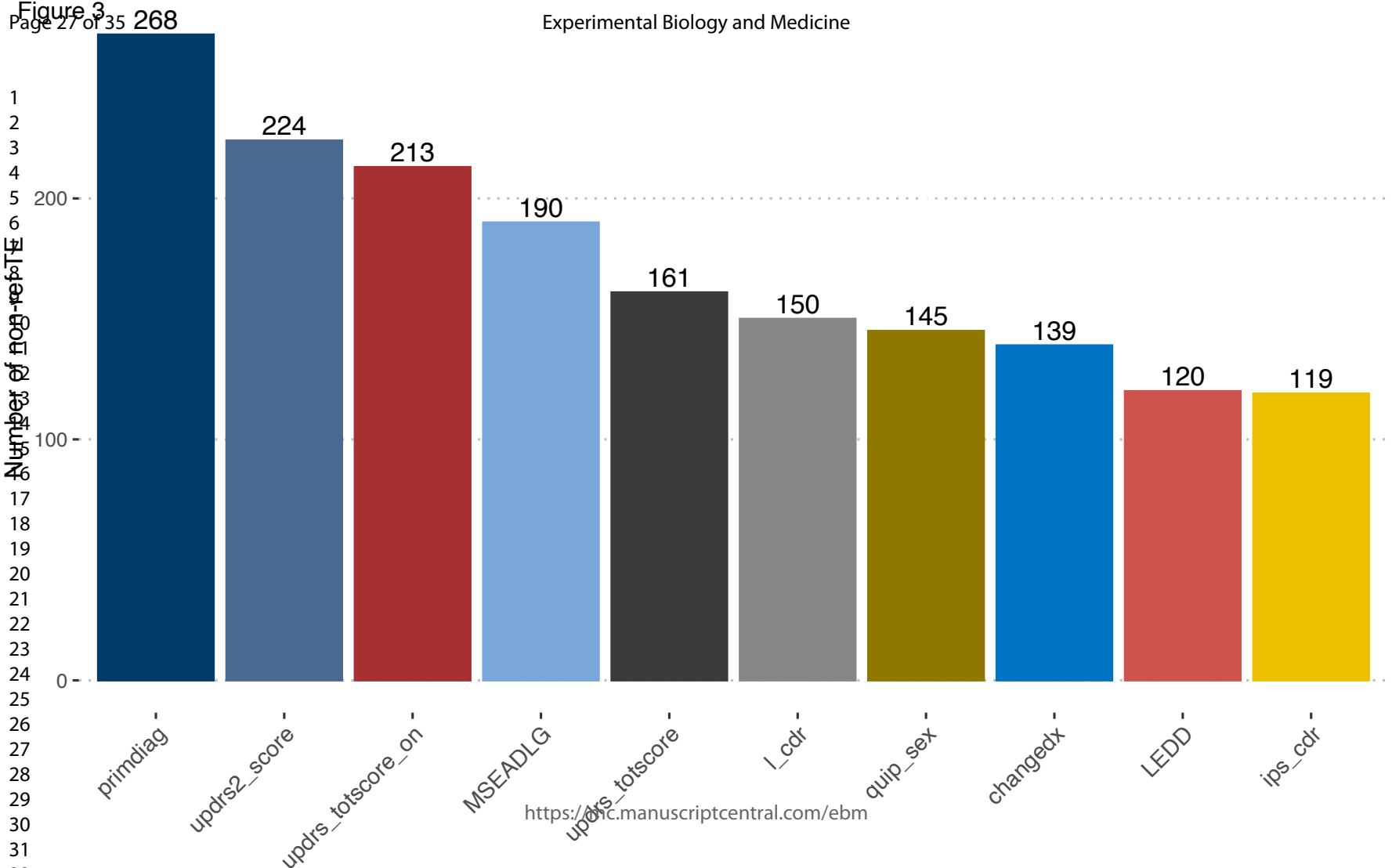
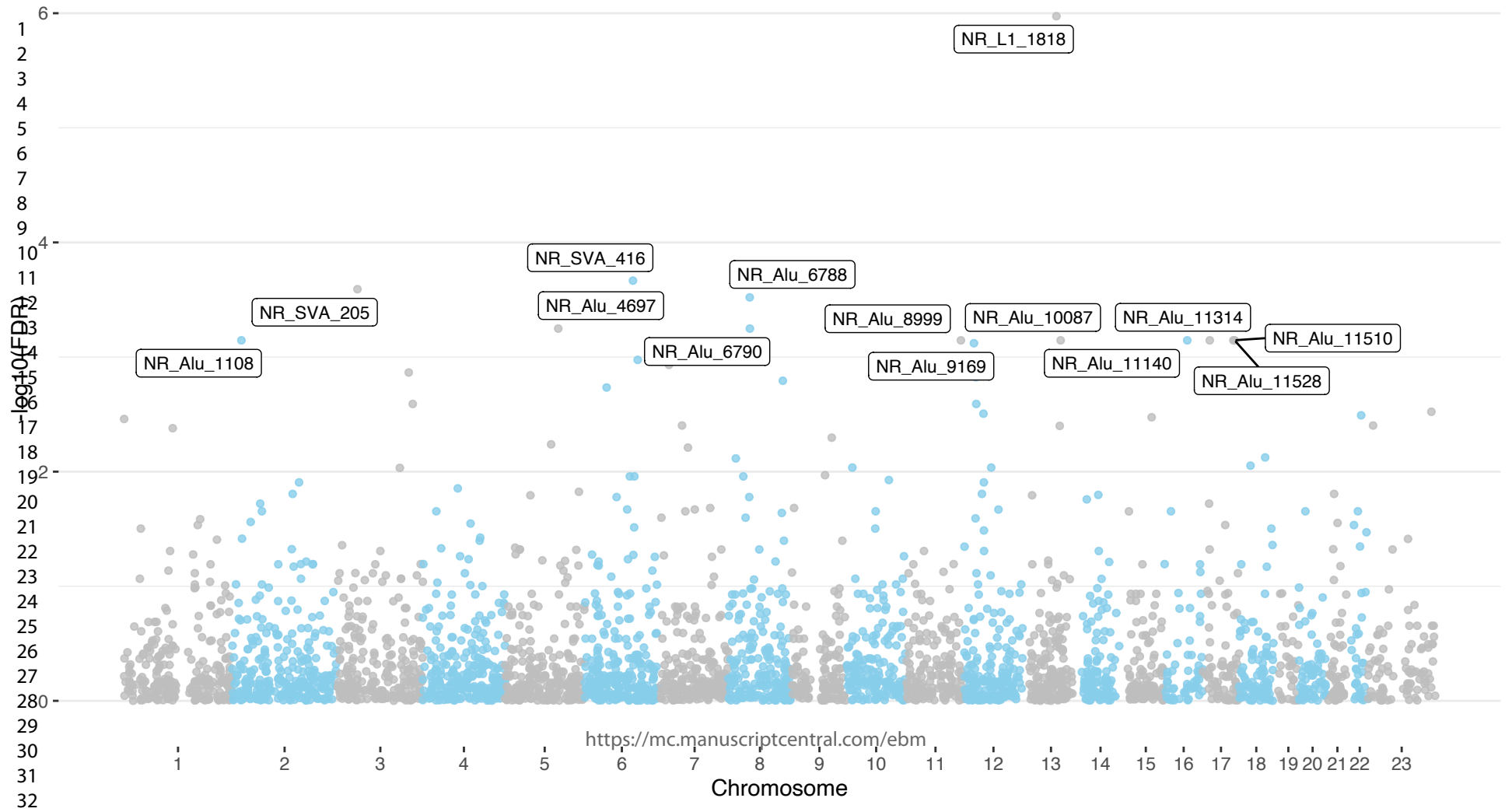

b
70
60
50
40
30
20
10
0
Percentage (%)
5'UTR
3'UTR
Exonic
Intronic
Promoter
Intergenic
Alu
L1
SVA

Figure 2

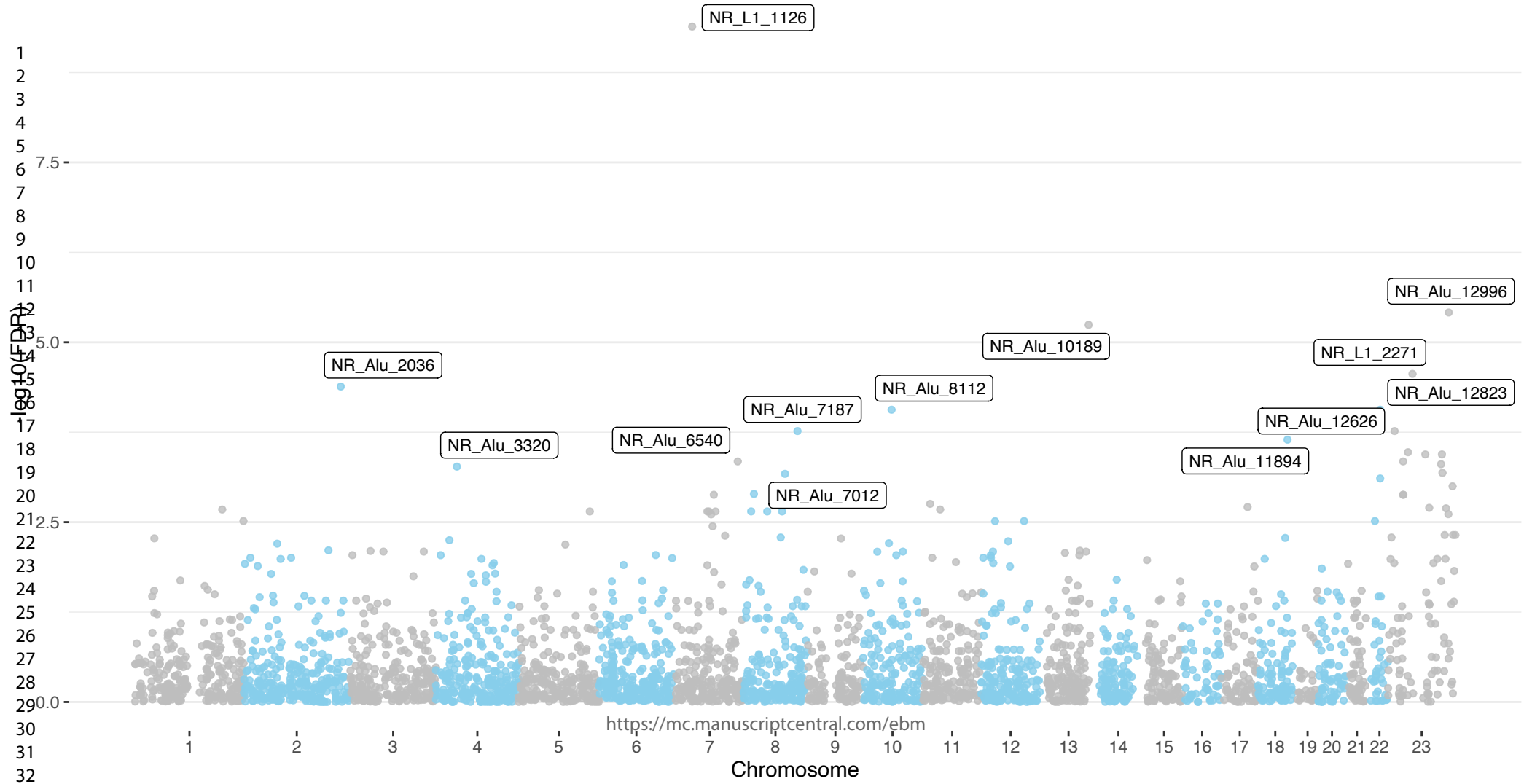

3000
2000
1000
0
2886
360
128
Alu
L1
SVA
type
Number of non-reference TEs

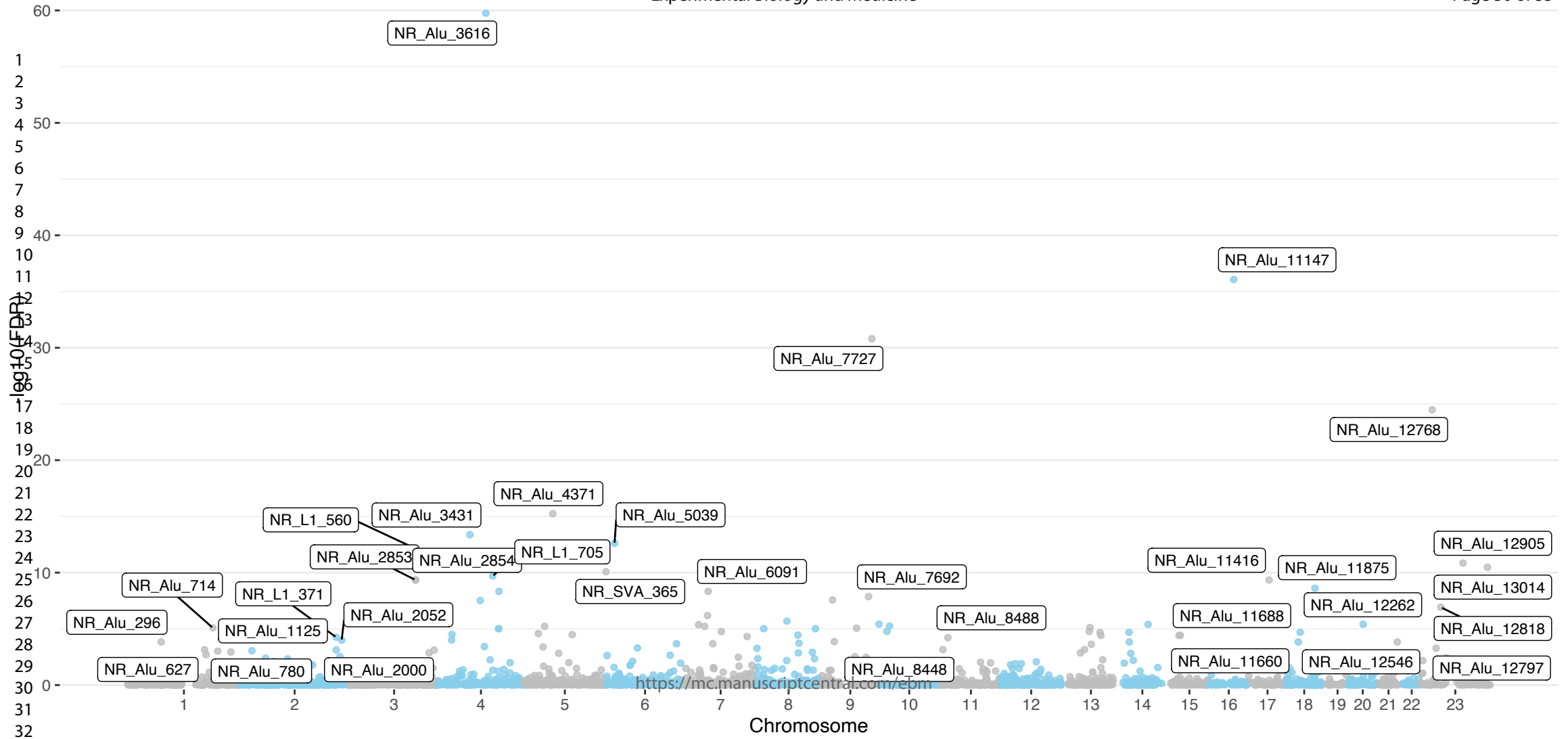

Number of non-redundant TE
0
100
200
268
224
213
190
161
150
145
139
120
119
primdiag
updrs2_score
updrs_totscore_on
MSEADLG
updrs_totscore
l_cdr
quip_sex
changedx
LEDD
ips_cdr

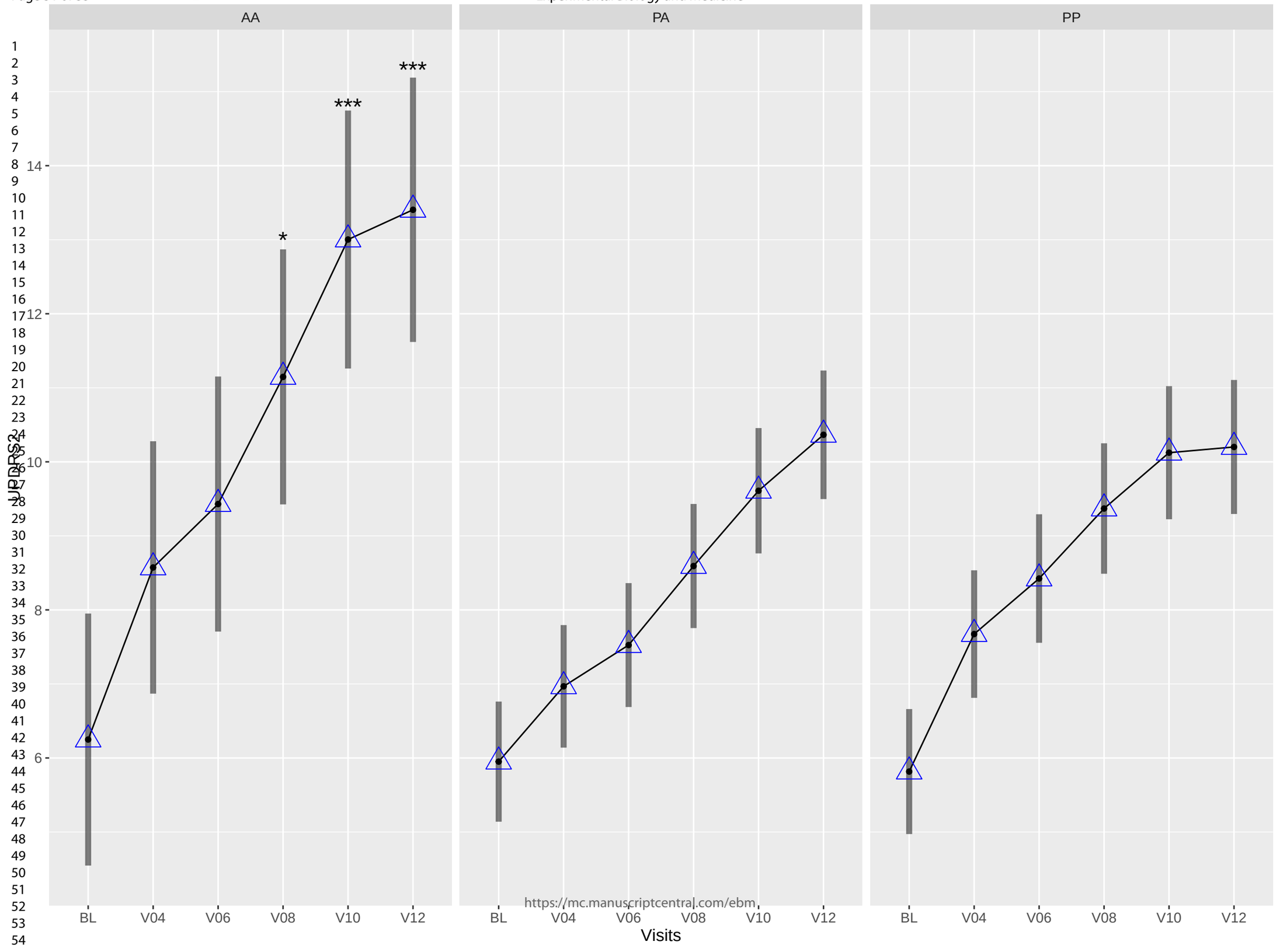

NR_L1_1818
NR_SVA_416
NR_Alu_6788
NR_Alu_4697
NR_SVA_205
NR_Alu_8999
NR_Alu_10087
NR_Alu_11314
NR_Alu_11510
NR_Alu_1108
NR_Alu_6790
NR_Alu_9169
NR_Alu_11140
NR_Alu_11528
−log10(FDR)
Chromosome
https://mc.manuscriptcentral.com/ebm

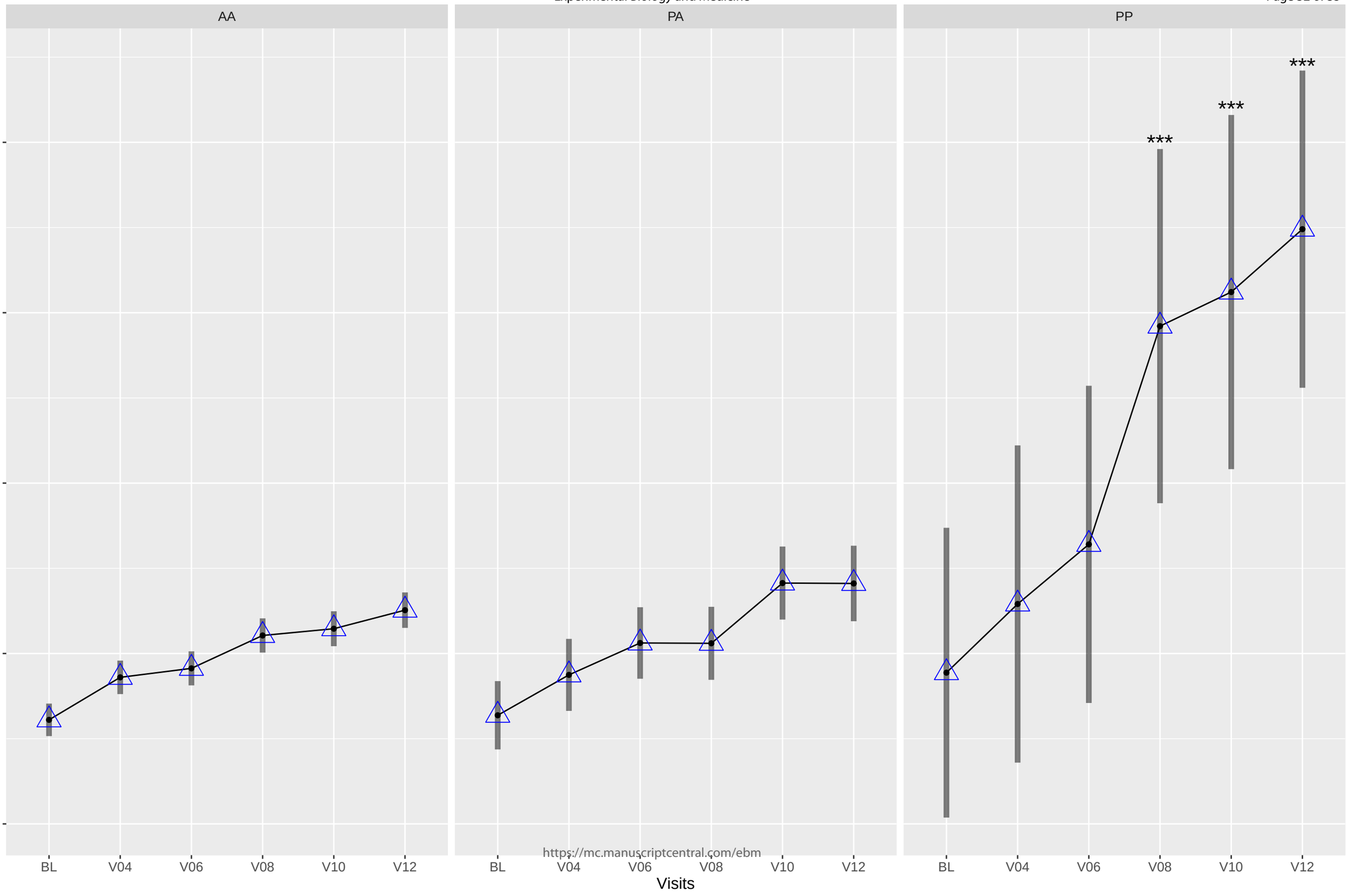
All prefTEs UPDRS2 PD Figure 5

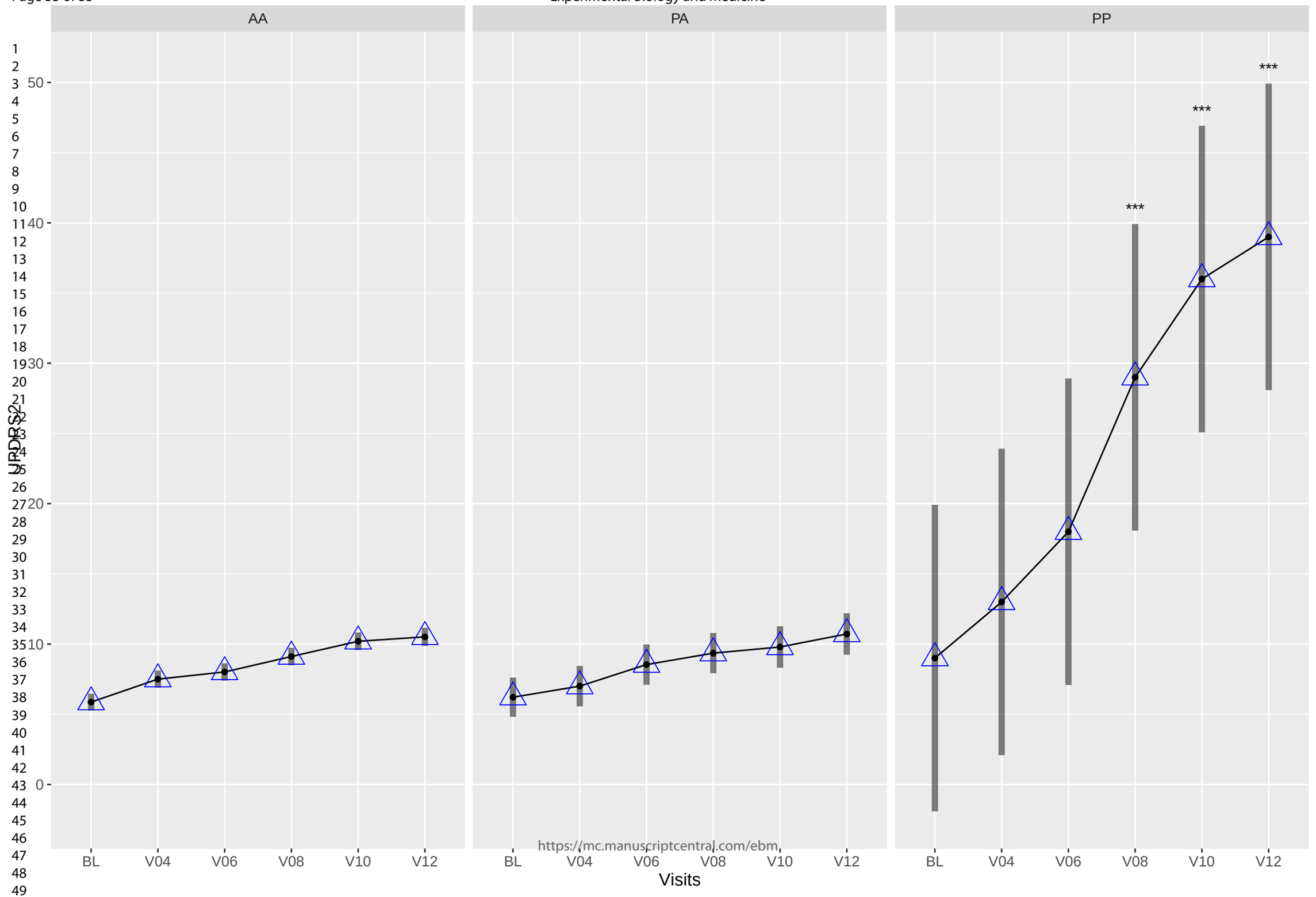
Experimental Biology and Medicine

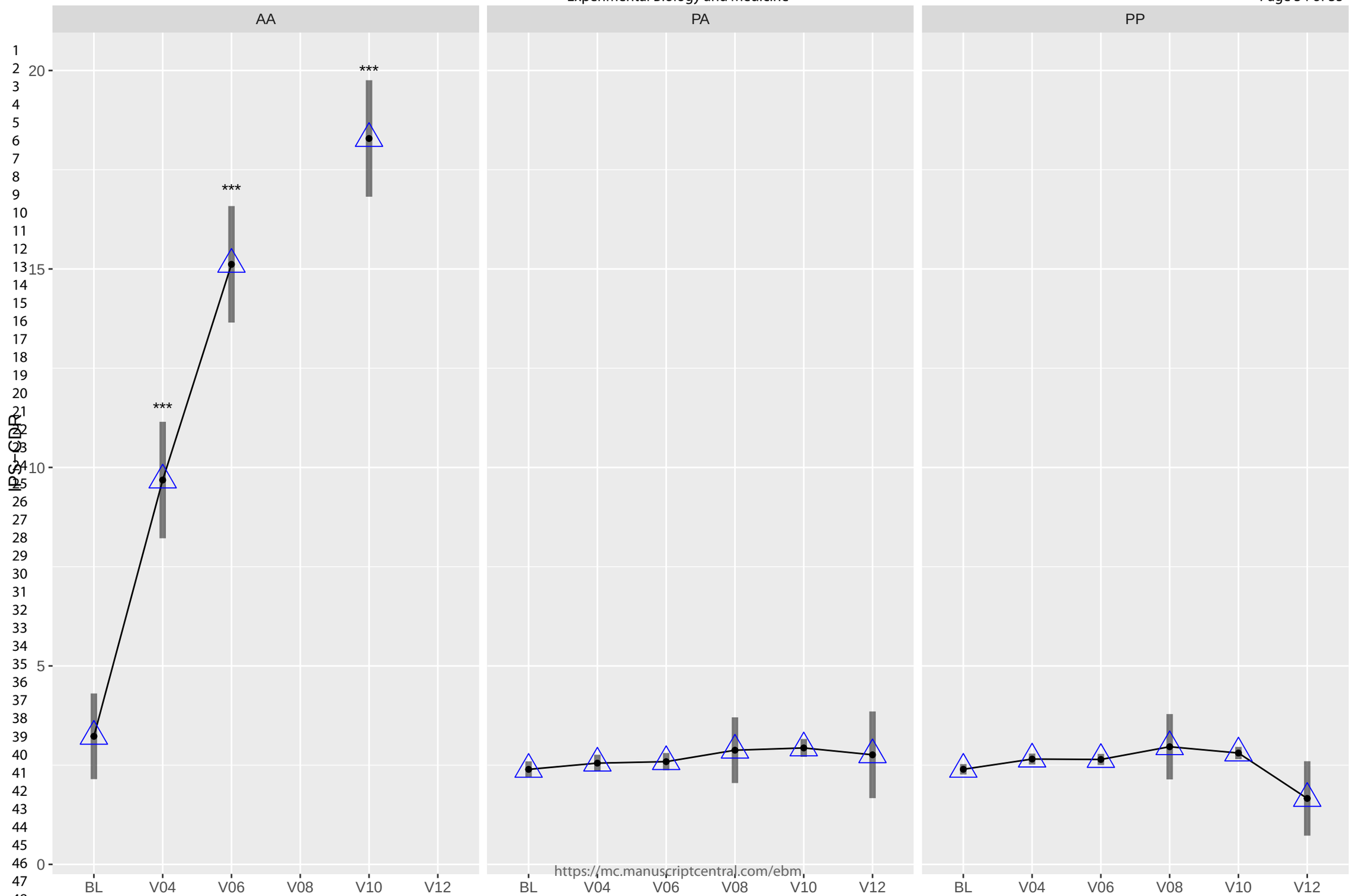


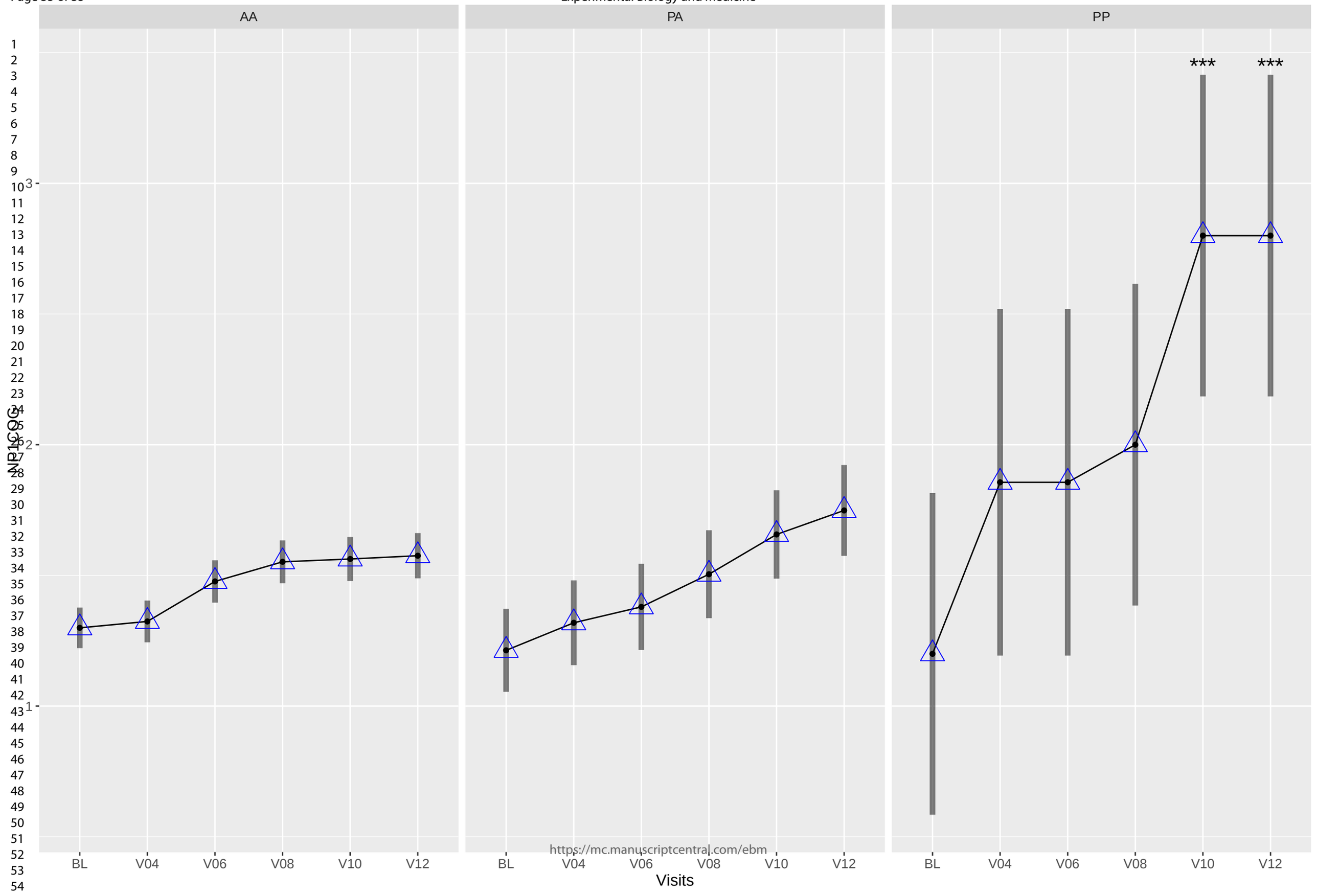

NR_Alu_3616
NR_Alu_11147
NR_Alu_7727
NR_Alu_12768
NR_Alu_4371
NR_L1_560
NR_Alu_3431
NR_Alu_5039
NR_Alu_2853
NR_Alu_2854
NR_L1_705
NR_Alu_6091
NR_Alu_7692
NR_Alu_11416
NR_Alu_11875
NR_Alu_12905
NR_Alu_714
NR_L1_371
NR_SVA_365
NR_Alu_13014
NR_Alu_296
NR_Alu_1125
NR_Alu_2052
NR_Alu_8488
NR_Alu_11688
NR_Alu_12262
NR_Alu_12818
NR_Alu_627
NR_Alu_780
NR_Alu_2000
NR_Alu_8448
NR_Alu_11660
NR_Alu_12546
NR_Alu_12797
-log10(FDR)
Chromosome
https://mc.manuscriptcentral.com/ebm

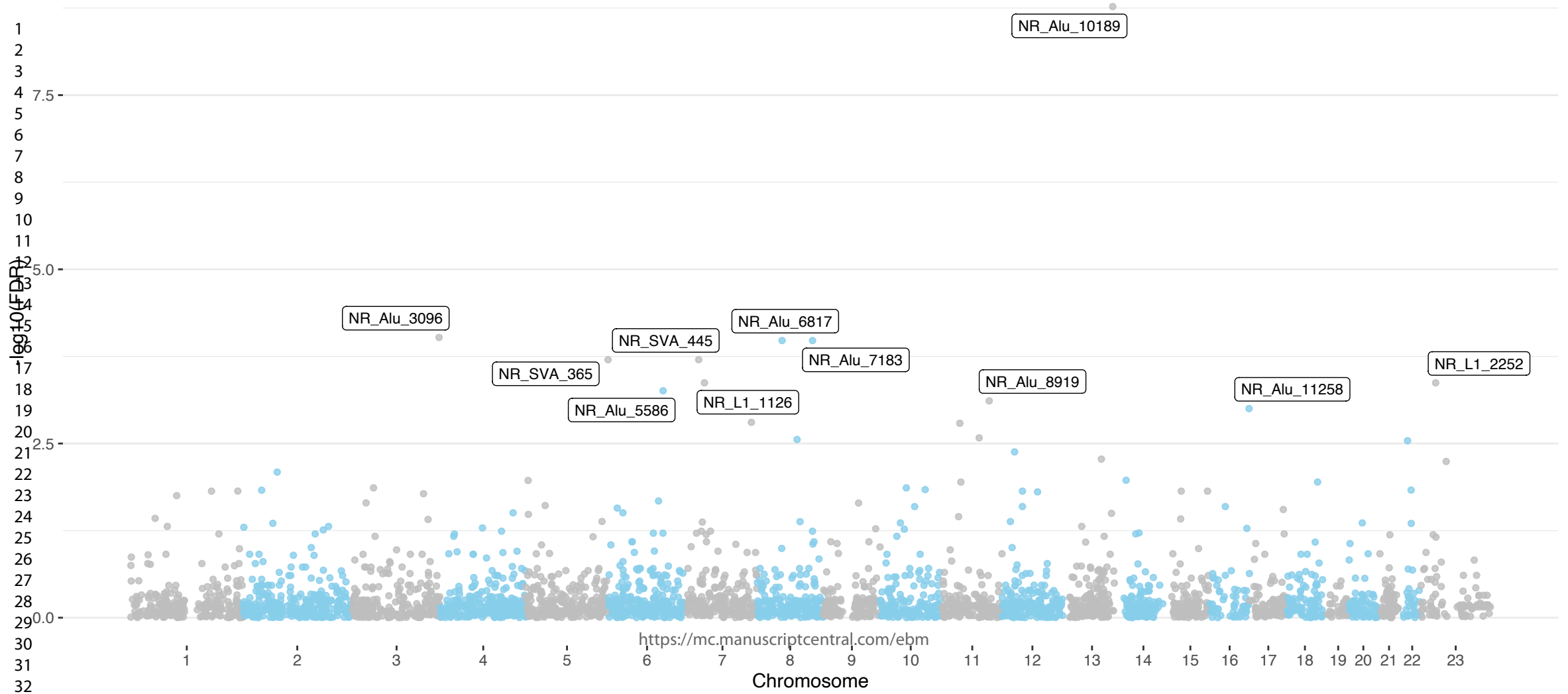

AA
PA
PP
*
***
***
14
12
10
8
6
UPDRS3
BL
V04
V06
V08
V10
V12
Visits


AA
PA
PP
UPDRS Total
100
80
60
40
20
***
***
***
BL
V04
V06
V08
V10
V12
Visits


AA
PA
PP
MDS-UPDRS2
50
40
30
20
10
0
***
***
***
BL
V04
V06
V08
V10
V12
Visits




AA
PA
PP
***
***
NRS (0-10)
3
2
1
BL
V04
V06
V08
V10
V12
Visits


NR_Alu_10189
NR_Alu_3096
NR_SVA_445
NR_Alu_6817
NR_SVA_365
NR_Alu_7183
NR_Alu_8919
NR_L1_2252
NR_Alu_11258
NR_Alu_5586
NR_L1_1126
7.5
5.0
2.5
0.0
-log10(FDR)
https://mc.manuscriptcentral.com/ebm
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
Chromosome