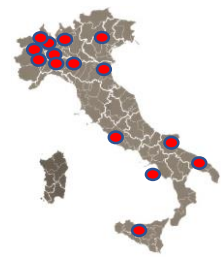
Integrating genetic and phenotypic data from the PROGEMUS database and the Italian Multiple Sclerosis and Related Disorders Register

D'Alfonso Sandra[1], Pomella Nicola[1,2], Lucisano Giuseppe[3], Barizzone Nadia[1], Thavamani Muralidharan[1], Carmela Pia Ferro[2], Trojano Maria[4], Leone Maurizio[2]

1.Department Of Health Sciences-CAAD – UPO University Of Eastern Piedmont – Novara – Italy 2.Scientific Research Department – IRCCS Casa Sollievo della Sofferenza – San Giovanni Rotondo (FG) – Italy 3.Department of Translational Biomedicine and Neurosciences–DiBraiN – University of Bari Aldo Moro – Bari – Italy 4.Department of Basic Medical Sciences, Neuroscience and Sense Organs – University of Bari Aldo Moro Policlinico – Bari – Italy

Objectives: The PROGEMUS Consortium was established in 2005 as a multi-centric project that enrolled over 3500 multiple sclerosis (MS) patients and collected genomic, clinical and environmental data for both cross-sectional and prospective follow-up study designs. The aim of this study is to implement, test and evaluate a procedure to link the PROGEMUS database with the Italian MS Register (RISM) [1], with the final goal to build an infrastructure of data sharing to explore possible associations of genetic variants with factors involved in clinical heterogeneity.



PROGEMUS Centers

RISM Centers

Materials and Methods: The PROGEMUS database was pre-processed and harmonized to discover potential outliers, data entry errors and lack of internal coherence. PROGEMUS MS centers were interrogated when needed to amend errors. Genetic data were collected by the PROGEMUS consortium, genotyped and imputed against the Haplotype Reference Consortium panel and the Four-Digit Multi-Ethnic HLA panel (figure 1).

The following polygenic risk scores (PRS) were calculated based on [2,3]:
-PRS_{201}:201 GWAS-significant, non-HLA susceptibility SNPs [2]
-$PRS_{201+HLA}$: 201 GWAS-significant, non-HLA susceptibility SNPs + 32 susceptibility HLA markers [2]
-$PRS_{500+HLA}$: 500 non-HLA susceptibility SNPs (GWAS-significant + suggestive) + 32 susceptibility HLA markers [2]
-$PRS_{severity}$: 10 non-HLA SNPs associated with MS severity [3]

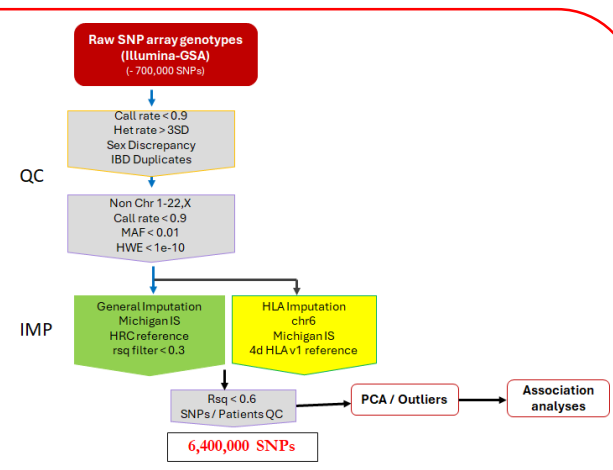


Figure 1. *Genetic analysis flowchart*

Linking fields between PROGEMUS and RISM were determined and pseudoanonimyzed; gender and date of birth in both databases were also used for quality control. We reached a final dataset of 1309 pwMS with comprehensive genetic, clinical and environmental data, inherited by both databases. Exploratory statistical analysis was performed for quality control, to evaluate concordance of data between the two databases and to determine associations between clinical and genetic factors.

Results and Discussion: The concordance rate varied between 53% and 94%, with the lowest values found for temporal variables - such as date of secondary progression, date of onset, date of second attack - whereas clinical outcomes such as type of onset, oligoclonal bands and others reached a very high concordance rate. Basic cross-sectional data analysis highlighted the significant impact of the $PRS_{500+HLA}$ susceptibility score on the number of functional systems at onset (monofocal vs. multifocal; p=0.0005, figure 2) and confirmed the association between HLA-DRB1*15:01 allele and the presence of oligoclonal bands in the CSF (p=0.0023, OR=4.40, 95%CI=1.33-14.57)

Regarding the known severity-associated variants, a non-HLA reported in [3] as GWAS-significantly associated with increase in ARMSS (rs10191329) is associated with a quicker Relapse Associated Worsening (RAW): p=0.002 after condition for SEX, comorbidities, multifocal onset, incidence of relapses (figure 3). Furthermore, we observed a trend for the $PRS_{severity:}$ pwMS with a lower risk score (quartile 4) for this PRS reach a PIRA (Progression Independent of Relapse Activity) event later than pwMS belonging to high risk categories, but the result is not statistically significant (figure 4).
Finally, we observed an effect on PIRA and CDA (confirmed disability accrual) for the HLA-A*02 allele (figure 5). Although HLA-A*02 is a known protective factor for MS susceptibility, it show a risk effect on disease severity, both on PIRA and CDA.

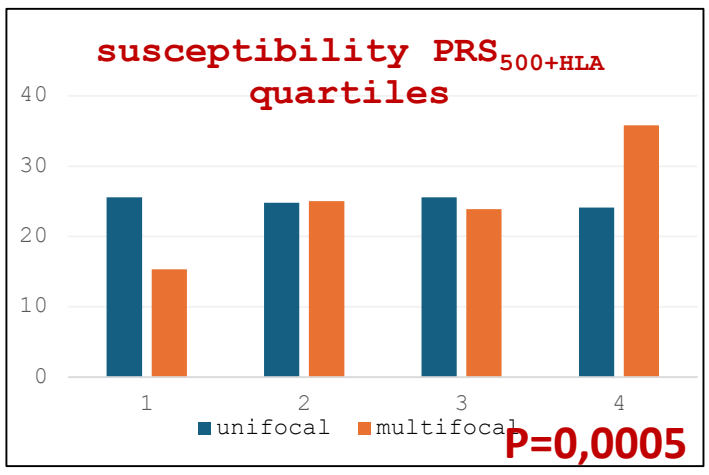


Figure 2. *pwMS have been divided in quartiles basing on susceptibility PRS value: 1: lowest risk, 4: highest risk.*

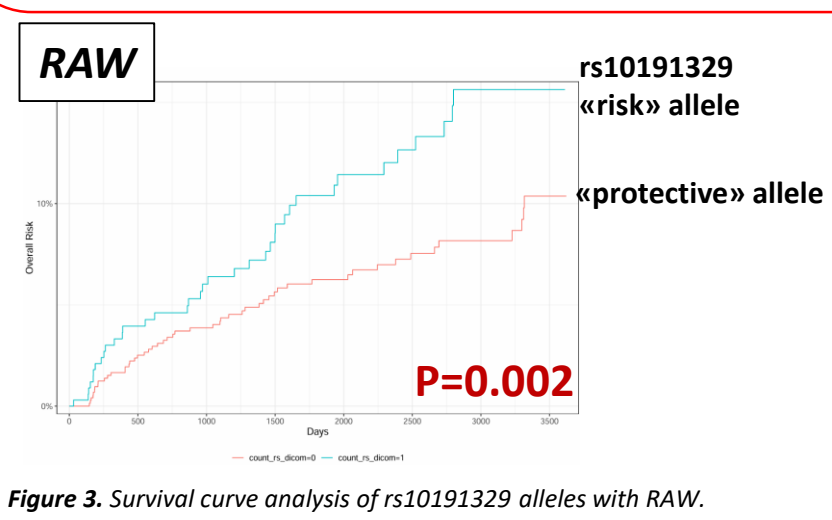


Figure 3. *Survival curve analysis of rs10191329 alleles with RAW. Univariated p-val=0.019. Logistic regression covaried by sex, comorbidities, multifocal onset, incidence of relapses: p-val=0.002.*

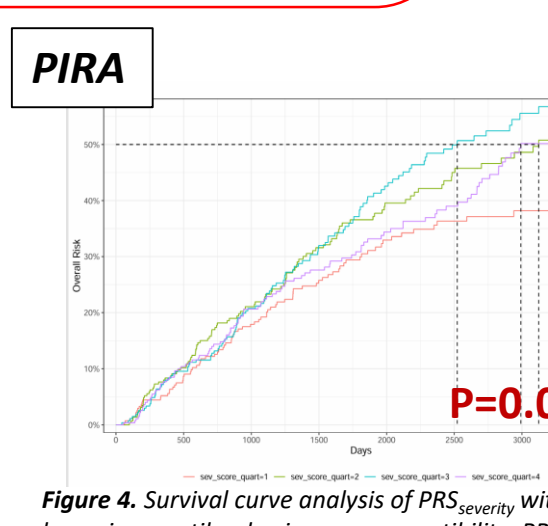


Figure 4. *Survival curve analysis of $PRS_{severity}$ with PIRA. pwMS have been in quartiles basing on susceptibility PRS value: 1: lowest risk, 4: highest risk*

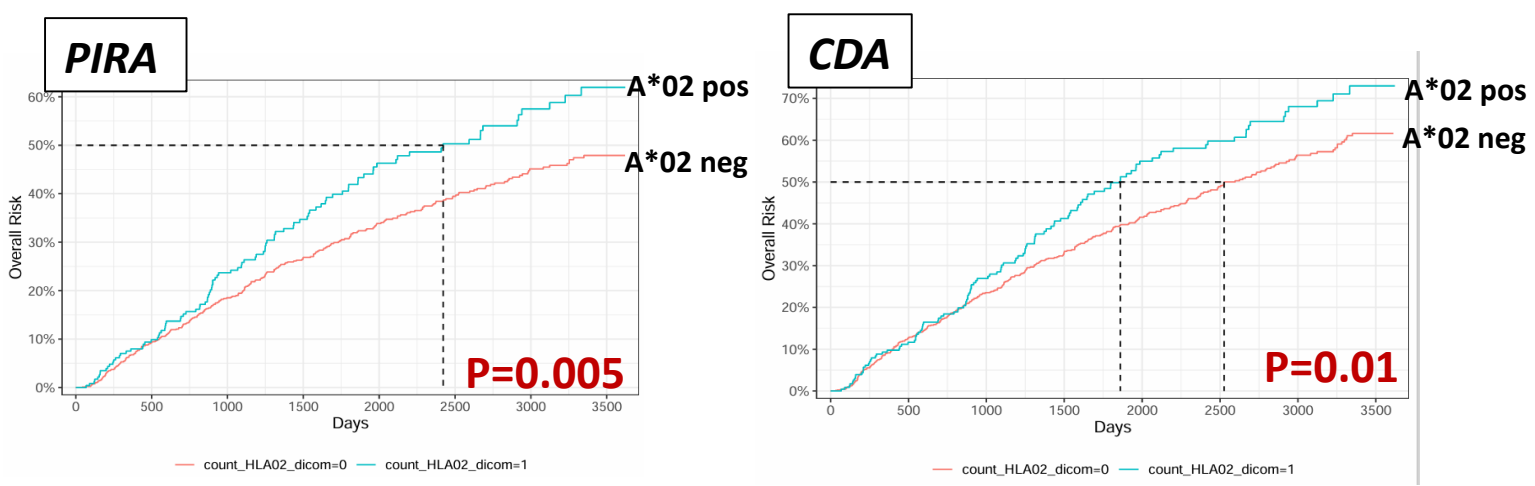


Figure 5. *Survival curve analysis of presence of HLA-A*02 allele with PIRA and CDA. People positive for the A*02 allele show a quicker worsening of disability.*

References

1)Trojano M, et al. The Italian multiple sclerosis register. Neurol Sci. 2019

2)International Multiple Sclerosis Genetics Consortium. Multiple sclerosis genomic map implicates peripheral immune cells and microglia in susceptibility. Science. 2019

3)International Multiple Sclerosis Genetics Consortium; MultipleMS Consortium. Locus for severity implicates CNS resilience in progression of multiple sclerosis. Nature. 2023

Conclusions: The linkage allows a high density of data, suitable to test complex outcomes (i.e. PIRA). There is an ongoing effort to further expand the number of linked patients and to populate the databases.