Beyond Genome wide association studies (GWAS) in multiple sclerosis (MS): fine mapping and functional analysis in MS susceptibility regions to identify novel drug target genes

Endri Visha[a], Alessandro Pizzino[a] , Ferdinando Clarelli[b], Lucia Corrado[a], Fjorilda Caushi[a], Beatrice Piola[a], Erica Melone[a], Diego Cotella[a], Laura Follia[a], Martina Tosi[a], , Domizia Vecchio[a], Massimo Filippi[c, d, e, f], Martinelli Boneschi Filippo Giovanni[g, h], Federica Esposito[b], [c],Maurizio Leone[i], Nadia Barizzone[a], Sandra D'alfonso[a]

BACKGROUND

Multiple Sclerosis (MS) is a complex autoimmune disease of the central nervous system caused by an interaction of multiple genetic and environmental factors (1) . Recently, international efforts (IMSGC 2011, 2013, 2019) (2,3,4) conducted through genome wide association studies (GWAS) have identified more than 30 HLA markers independent from the HLA-DRB1*15:01 and 201 non HLA MS risk loci, one of these located on the X-chromosome . However, for most of them, the specific functional mechanism by which these loci influence MS pathogenesis is still largely unknown. Moreover, most MS susceptibility regions contains several genes and many variants and the primarily associated ones are still unknown.
GWAS studies still carry some intrinsic limits represented by linkage disequilibrium (LD) inability of pinpointing the real causative variant among associated SNPs. Fine mapping studies have the objective of overcoming these limitations, possibly highlighting new pathogenic mechanisms which could be the target for new therapeutic approaches. Pinpointing causal variants in MS associated loci containing drug target genes (genes encoding for proteins targeted by an already approved drug or by a molecule currently tested in clinical trials), could potentially lead to repurposing of already known drugs, whose molecular target may overlap with genes involved in MS pathogenesis .

AIM

We aimed to fine map and funcrtionally analyse MS susceptibility loci to identify the functional mechanism by which they influence the disease pathogenesis. To enhance the translational potential, among MS regions we focused on those containing genes which are targets of already existing drugs (drug target genes).

MATERIALS AND METHODS

Region selection: We performed an analysis of co-occurrence of drug target genes with MS susceptibility regions using three databases (Drug Bank, Therapeutic Target Database (TTD), ChEMBL) and identified an overlap of 156 genes (93 regions) .
Among these, we focused on 11 regions (containing 23 drug target genes) because they are significantly associated on our large Italian cohort consisting of 4530 MS patients and 3482 Healthy Controls with GWAS data (**Figure 1**), genotyped with different array platforms.
Figure 2 shows an example of regional association plot on the Italian sample set for one of the selected regions.

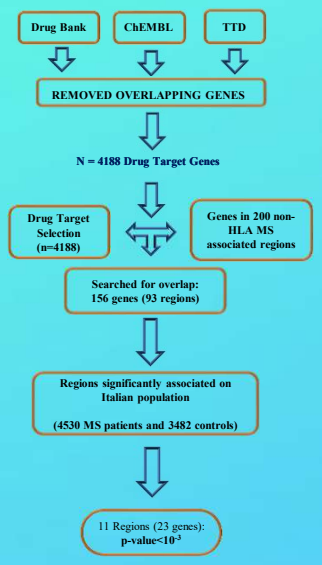

Figure 1: Flowchart for drug target selection in MS regions

In-silico functional prediction: For these 11 regions we integrated GWAS results with data from in silico resources, like eQTL data (Phenoscanner) and differential expression data between MS patients and healthy controls obtained through two array expression data sets collected in GEO (*Gene Expression Omnibus*) and thus selected for further investigation 5 regions where the lead SNP is a putative eQTL SNP for a drug target gene.

To experimentally verify this hypothesis and to functionally analyse the lead SNP and all the variants in strong LD with it, we used Massively Parallel Reporter Assay (MPRA), a high-throughput screening method able to test thousands of sequences for their putative transcription regulation role, to identify variants that modulate gene expression.

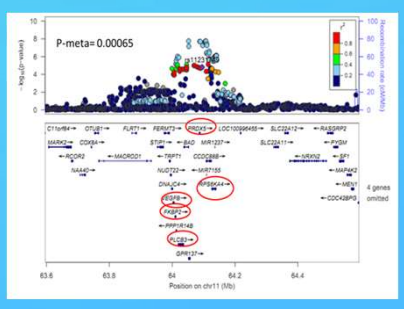

Figure 2 : Regional plot[2] of region containing rs11231749. This is the regional association plot of one of the selected regions, showing a strong association in The genes highlighted in red are drug-target genes. The SNP with the purple diamond is the most strongly associated in the international cohort. The reported p-value is from the meta-analysis on the Italian cohort.

MPRA methods

1. **Oligonucleotide Library Synthesis.(Fig.3)**
Each SNP is represented by 60 oligos based on its 145 bp sequence 30 oligos are designed on the forward and 30 on the reverse strand. For each strand, 10 oligos represent the reference allele, 10 the alternative allele and 10 the scrambled allele, used as a control. Primer1seq and Primer2seq are fixed sequences needed for PCR amplification of the library. GGTACC is the sequence recognized by KpnI restriction enzyme and TCTAGA is the one cut by XbaI. 10 bp Tag distinguishes each oligonucleotide.

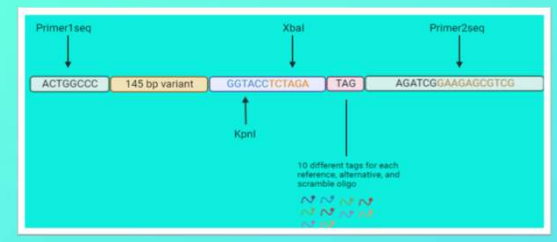

Figure.3

2. ePCR. Aimed to the amplification of the library and the addition of SfiI restriction sites (GGCCNNNNNGGCC) at the extremities of the library needed for the following first cloning step (Fig.4)

Figure.4

3. First cloning step. Enzymatic digestion of pMPRA1 backbone plasmid and amplified MPRA library with SfiI restriction enzyme in order to allow the ligation between them. The produced intrmediate plasmid library is then used to transform DH5α electrocompetent bacterial cells.(Fig.5)

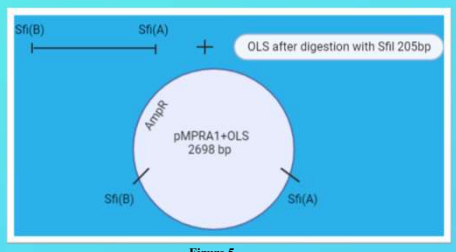

Figure.5

4. Second cloning step. Enzymatic digestion of the intermediate plasmid library and pMPRAdonor1 or pMPRAdonor2 plasmids with KpnI and XbaI restriction enzymes in order to allow the ligation between the intermediate plasmid library and Luc2 ORF coming from pMPRAdonor1 digestion (to test the possible promoter role of the selected SNPs), or between intermediate plasmid library and a minimal promoter with Luc2 ORF coming from pMPRAdonor2 (to test the possible enhancer role of the selected SNPs).(Fig.6)

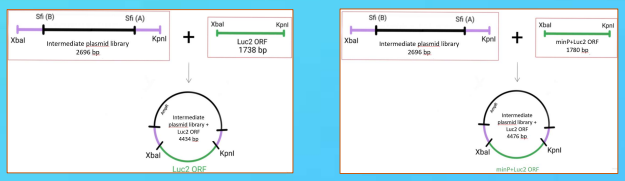

Figure.6

5. **Transfection.** The resulting plasmid libraries are transfected into SH-SY5Y neuroblastoma derived eukaryotic cells through jetprime® polyplus transfection reagents.

6. **Tag-sequencing analysis**. Based on the amount of plasmid DNA starting material and its consequent produced RNA, information on transcriptional regulation impact of the tested variants can be derived by comparing RNA/DNA tag count ratio (obtained through NGS) between reference, alternative, and scramble (negative control) alleles.

RESULTS

The co-occurrence analysis and the successive in silico analysis allowed us to prioritize 5 drug target genes, located in regions associated with MS in the Italian population, to be further analysed with the MPRA (Table 1).

Table 1. Genes selected for MPRA analysis

Gene[a]	PeakSNP[b]	CHR[c]	Pvalue[d]	OR[d]	Drug name[e]	eQTL effect[f]
TEC	rs6837324	4	0.00090	1.19	Fostamatinib	risk allele > expression
TXK					Fostamatinib	risk allele < expression
PRDX5	rs11231749	11	0.00065	1.20	Auranofin	risk allele < expression
VEGFB					Aflibercept	risk allele < expression
CD40	rs6032662	20	0.0014	1.18	ISIS 19211	risk allele < expression
IFNGR2	rs9808753	21	0.0042	1.26	Interferon gamma-1b	risk allele > expression

A total of 86 SNPs within the five selected regions, including the 5 lead SNPs and variants in strong LD ($r^2 \geq 0.77$), were selected to create a chip for MPRA analysis, constructing a plasmid containing a library composed of 5000 probes. We could assess the enhancer function of each variant by comparing RNA/DNA tag count ratio obtained through NGS and analysed by the MpraLM tool.
For each of the 5 selected regions, preliminary data identified at least one SNP influencing gene expression of a drug target gene, with a statistically significant effect.

CONCLUSIONS

These promising results need further validation with in silico and in vitro approaches, and have potential implication on the knowledge of disease mechanisms and novel drug target. This approach can be applied to other diseases Indeed, literature studies demonstrated that pipeline drug targets with genetic evidence of disease association are twice as likely to lead to be approved .

UPO AGING PROJECT
UNIVERSITÀ DEL PIEMONTE ORIENTALE

1. Compston Alastair, Alasdair Coles. Multiple sclerosis. Lancet. 2008 Oct 25;372(9648):1502-17
2. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, Boehnke M, Abecasis GR, Willer CJ. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics. 2010 Sep 15;26(18):2336-7. doi: 10.1093/bioinformatics/btq419. Epub 2010 Jul 15. PMID: 20634204; PMCID: PMC2935401.