

**Revisiting the link between platelets and depression through genetic epidemiology: new insights from platelet distribution width**

---

*Alessandro Gialluisi,<sup>1</sup> Benedetta Izzi,<sup>1</sup> Augusto Di Castelnuovo,<sup>2</sup> Chiara Cerletti,<sup>1</sup> Maria Bendetta Donati,<sup>1</sup> Giovanni de Gaetano<sup>1</sup> and Licia Iacoviello<sup>1,3</sup>*

*<sup>1</sup>Department of Epidemiology and Prevention, IRCCS Neuromed, Pozzilli; <sup>2</sup>Mediterranea Cardiocentro, Napoli and <sup>3</sup>Department of Medicine and Surgery, University of Insubria, Varese, Italy*

*Correspondence: ALESSANDRO GIALLUISI - [alessandro.gialluisi@gmail.com](mailto:alessandro.gialluisi@gmail.com)*

*doi:10.3324/haematol.2019.222513*

## **Supplementary Methods**

### **LD score regression analysis**

LD score regression models genetic correlation between two (either continuous or binary) traits as a function of LD (linkage disequilibrium) score among SNPs in 1 cM bins genome-wide, as reported below:<sup>1,2</sup>

$$r_g = \rho_g / \sqrt{h_1^2 * h_2^2},$$

where  $\rho_g$  is the genetic covariance between trait 1 and trait 2, and  $h_1^2$  and  $h_2^2$  represent the SNP-based heritability of the two traits. SNP-based heritability is in turn computed as the slope of the linear function between  $\chi^2$  association statistics and LD score (i.e. the sum of  $r^2$  of a given SNP with all the other SNPs in a 1 cM window), for every SNP tested genome-wide (i.e. for which the association statistics are available in a given GWAS study).

In our analysis, we used only common SNPs which were available in the HapMap 3 reference panel<sup>3</sup> -excluding the HLA region- since these variants have good imputation quality stats ( $r^2 > 0.9$ ) in most studies.<sup>2</sup> LD scores of these variants were derived using the 1000G phase 1 v3 EUR panel (available at <https://data.broadinstitute.org/alkesgroup/LDSCORE/>). A Bonferroni corrected significance threshold was set to  $\alpha = 0.017$ , taking into account three platelet parameters analysed.

### **Genome-wide Mendelian Randomization analysis**

Mendelian Randomization is a technique which allows to infer causality relationships between two phenotypes – one hypothesized to be the exposure and one to be the outcome - through their associations with selected genetic variants, or instruments. This method is based on three basic assumptions: i) the genetic variant is associated with the exposure; ii) the genetic variant is not associated with confounders of the relation between exposure and outcome; and iii) the genetic variant influences the outcome only through its influence on the exposure. In the last years, methods have been published which allow to carry out MR at the genome-wide level, exploiting summary association statistics of GWAS

publicly available. For the purpose of this study, we used for our MR analyses the R version of the MRbase online tool (<http://www.mrbase.org/>)<sup>4</sup>, known as TwoSampleMR package (v 0.4.2). We used different alternative methods for this analysis, to have robust estimates (see Table 2a, b in the main text of the manuscript, and ref. <sup>4</sup> for a complete list). In particular, we focused our attention on Inverse-variance weighted (IVW) linear regression, which models the relation between exposure and outcome by interpolating the SNP-exposure and SNP-outcome effect for each instrumental variant analysed, weighting the contribution of each instrumental SNP to the overall effect by the inverse of the variance of the SNP-outcome effect.<sup>4</sup> Also, we checked the robustness of our results through Egger regression, which models the relationship between exposure and outcome as in IVW regression, but does not constraint the intercept to zero, so returning an unbiased effect estimate even in presence of unbalanced horizontal pleiotropy. This reflects the situation where the instrumental variants show non-relational pleiotropic effects on both outcome and exposure which significantly differ from zero, which may be due to the influence of unknown pathways on the outcome which are not dependent on the exposure.<sup>5</sup>

Under the hypothesis of bi-directional causality, we modelled the MR regression assuming PDW as exposure and MDD as outcome, and viceversa. For each analysis, instrumental variables were selected which showed genome-wide significant associations with the exposure in the original work ( $p\text{-value} < 5 \times 10^{-8}$ ), removing palindromic SNPs and applying a strict LD clumping to ensure independence among SNPs ( $r^2$  cutoff=0.001 and clumping window=10,000 kb). After filtering, regressions were modelled on 114 SNPs and 4 SNPs meeting these criteria, when assuming PDW and MDD as exposure, respectively.

Study	Phenotype	Ncases <sup>a</sup>	Nctrl <sup>a</sup>	Ntot <sup>a</sup>	URL <sup>b</sup>
Wray et al., 2018 <sup>6</sup>	MDD	59,851	113,154	173,005	<a href="https://www.med.unc.edu/pgc/results-and-downloads">https://www.med.unc.edu/pgc/results-and-downloads</a>
Astle et al., 2016 <sup>7</sup>	Plt	NA	NA	166,066	<a href="ftp://ftp.sanger.ac.uk/pub/project/humgen/summary_statistics/human/2017-12-12/">ftp://ftp.sanger.ac.uk/pub/project/humgen/summary_statistics/human/2017-12-12/</a>
	MPV	NA	NA	164,454	
	PDW	NA	NA	164,433	

**Table S1.** GWAS studies which were used for the LD score regression and Mendelian Randomization analyses.

<sup>a</sup> These figures refer to the sample size for which GWAS meta-analysis results were actually available for download. In the case of Wray et al<sup>6</sup>, this was lower than the totality of samples analysed in the original study, due to restrictions of public access to data.

<sup>b</sup> URLs where association statistics are available for download. Number of cases and controls are reported where applicable (MDD case-control GWAS study). Abbreviations: MDD = major depressive disorder; Plt = platelet count; MPV = mean platelet volume; PDW = platelet distribution width.

## References

1. Bulik-Sullivan B, Loh PR, Finucane HK, et al. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet.* 2015;47(3):291-295. doi:10.1038/ng.3211.
2. Bulik-Sullivan B, Finucane HK, Anttila V, et al. An atlas of genetic correlations across human diseases and traits. *Nat Genet.* 2015;47(11):1236-1241. doi:10.1038/ng.3406.
3. The International HapMap C. Integrating common and rare genetic variation in diverse human populations. *Nature.* 2010;467:52. doi:10.1038/nature09298<https://www.nature.com/articles/nature09298#supplementary-information>.
4. Hemani G, Zheng J, Elsworth B, et al. The MR-base platform supports systematic causal inference across the human phenome. *Elife.* 2018;7:1-29. doi:10.7554/eLife.34408.
5. Burgess S, Thompson SG. Interpreting findings from Mendelian randomization using the MR-Egger method. *Eur J Epidemiol.* 2017;32(5):377-389. doi:10.1007/s10654-017-0255-x.
6. Wray NR, Ripke S, Mattheisen M, et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat Genet.* 2018;50(5):668-681. doi:10.1038/s41588-018-0090-3.
7. Astle WJ, Elding H, Jiang T, et al. The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common Complex Disease. *Cell.* 2016;167(5):1415-1429.e19. doi:10.1016/j.cell.2016.10.042.