Single cell sequencing data obtained and processing. The single-cell data were downloaded from NCBI website GSE211630 which had oral mucosa samples from one healthy volunteer, three non-erosive OLP and two erosive patients [6]. Next, we carried out data quality control and captured cells with less than 10% of mitochondrial genes, with a total number of genes ranging from 200 to 8000 and were expressed in at least three cells. The number of highly variable genes was set at 2000. The seven samples with one healthy control were integrated through the harmony algorithm. Then, UMAP method was used to reduce the dimension of data. The single cell-RNA sequencing method was used to map the hub genes and locate their cell source. CellChat R package (1.4.0) and slingshot package (2.10.0) were applied to explore the cell-cell interaction between the cell clusters.

Pseudo-time analysis. The R package Monocle v2.22.0 was applied for pseudo-time analysis to obtain the genes required for calculating differential gene expression from different pathological stages. After calculating the pseudo-time, the differential gene expression analysis was repeated to determine the genes that changed as a function of pseudo-time. The cell state containing the greatest number of S0-stage cells was considered as the root state. The threshold of the q value for multiple testing involved in the selection of differently-expressed genes (DEGs) was 0.01.

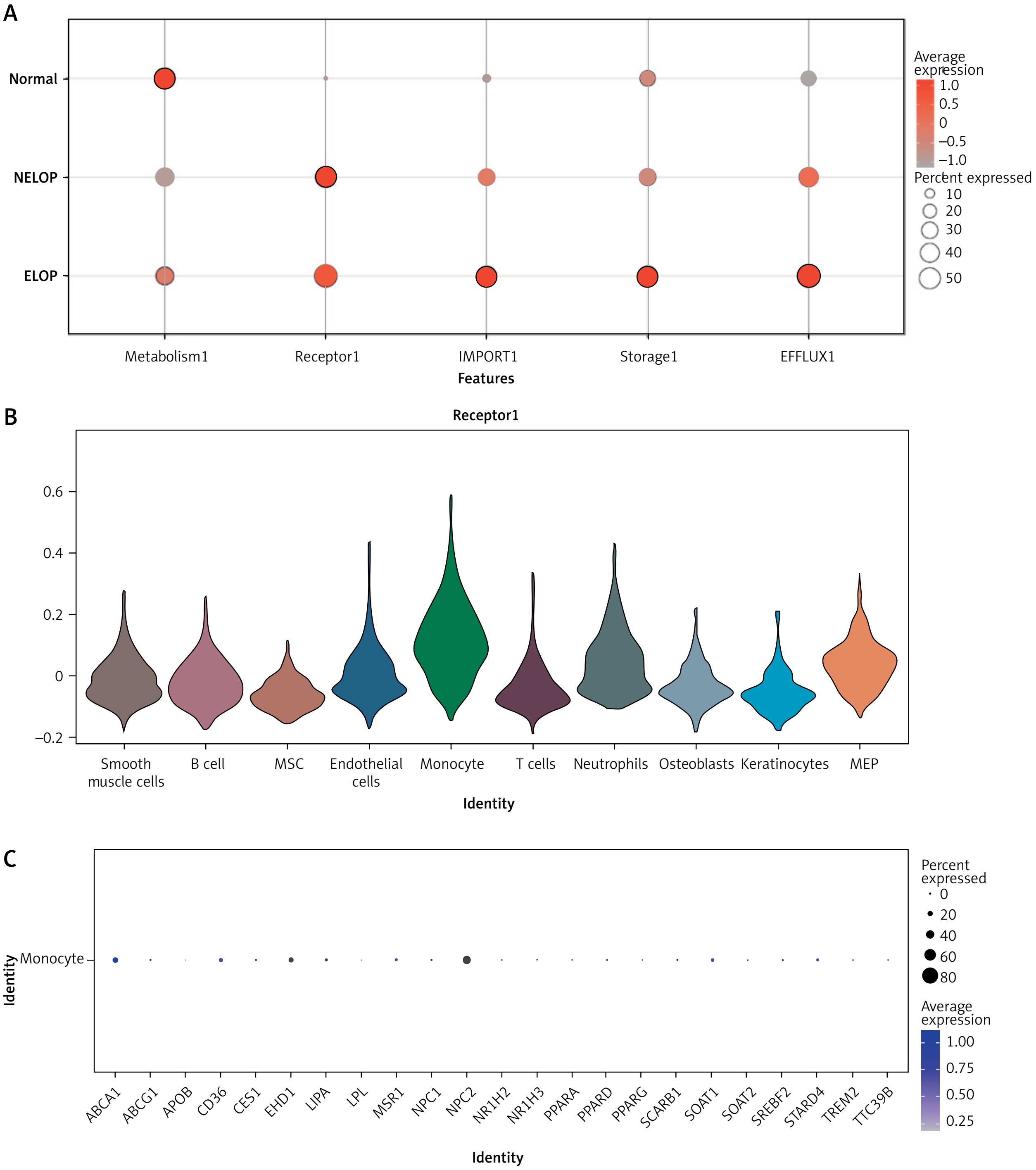
Study design and data sources for MR. According to previous studies, we designed two-sample MR methods to estimate the causal effects of plasma metabolites on OLP using genetic predictors of specific metabolite expression quantitative trait loci (eQTL) as instrumental variables (IVs). Single nucleotide polymorphisms (SNPs) as IVs for each trait were extracted from previously published GWAS and clumped together to obtain independent genetic variants (linkage disequilibrium (LD) $r^2 < 0.001$; clumping distance of 1000 kb). An F-statistic larger than ten was considered strong enough to predict the exposure of interest. Datasets based on ethnic groups of European ancestry were used to reduce the selection bias and improve the robustness of the analysis.

Data for outcome. We used FinnGen data on OLP from contributing studies including 510/376767 participants of European ancestry (K11_ORAL_LICHEN_PLANUS). The coefficient for each SNP was transformed into log odds ratio (OR).

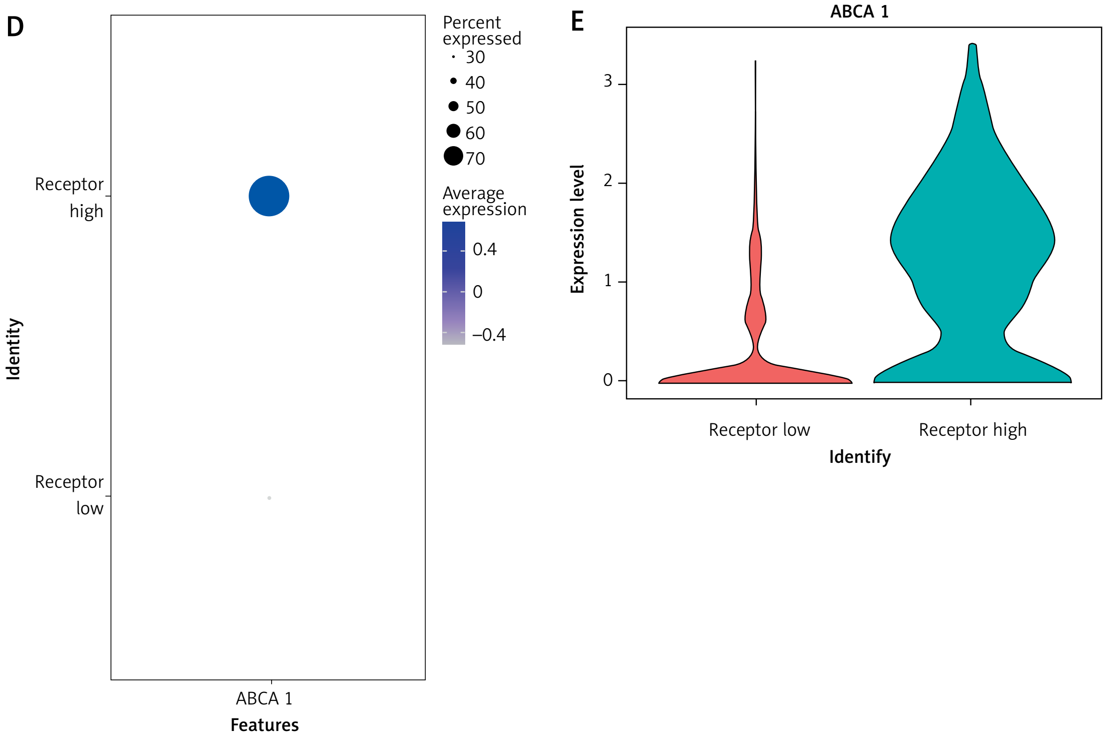
Statistical analysis. We used the two-sample MR method to evaluate the causal effects of eQTL of hub genes on outcomes and presented the results as OR and 95% confidence intervals (CIs). Harmonization of SNP exposure and outcomes has been previously described. Five MR approaches (random-effect IVW, weighted median, MR-Egger, simple mode, and weighted mode) were implemented to address the possible link between hub genes and OLP. The IVW regression approach assumes that all genetic variants are valid instruments or that there is no evidence of the pleiotropy effect. The weighted median was implemented to assess the robustness of the consistent estimation if more than 50% of the IVs were valid. We performed MR-Egger analysis using weighted linear regression between SNP exposure and SNP outcome. We detected the presence of potentially horizontal pleiotropy by the intercept of the MR-Egger coefficient. Heterogeneity statistics were performed using scatter plots between causal estimates from multiple genetic variants.

Sensitivity analysis. In the sensitivity analyses, we compared the causal estimates from various MR methods using MR-Egger, penalized weighted median, simple mode, IVW, and weighted mode to improve the robustness of the results. Forest plots were used to assess the causal effects using each of the SNPs on their own and to further compare against the causal estimates using the IVW and MR-Egger approaches that used all enrolled SNPs. We also examined possible directional pleiotropy by observing asymmetry in the funnel plots to gauge the reliability of the current MR analyses. Each time, leave-one-out sensitivity analyses were performed by removing a single SNP to assess whether the variant drove the association between exposure and the outcome variable.

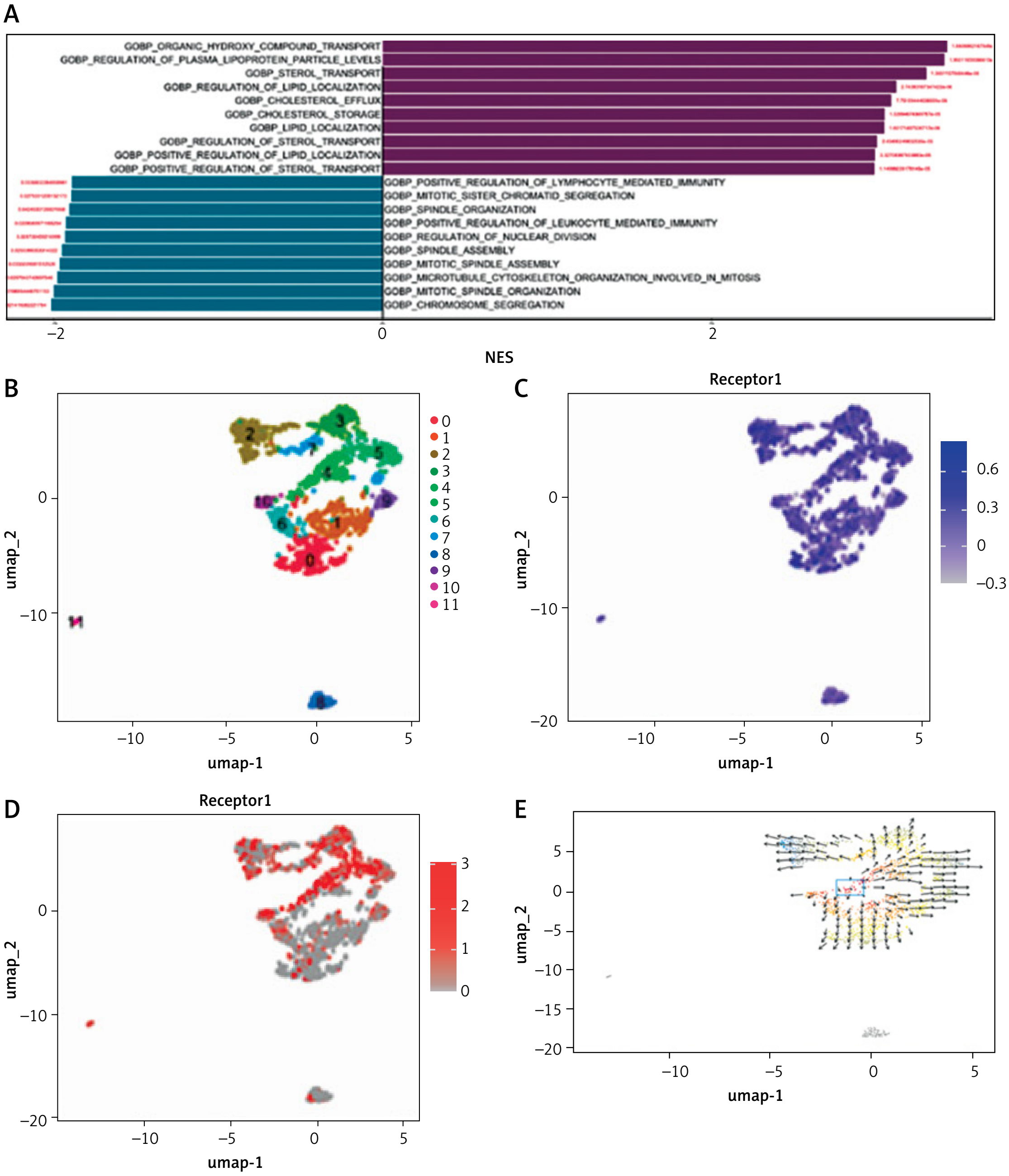
Validation with MR-IVW. For causal estimates from MR studies to be valid, three assumptions must be met: (1) the genetic variants are strongly associated with the exposure, (2) the genetic variants are not associated with any potential confounder of the exposure–outcome association and (3) the variants do not affect outcome independently of exposure. To meet the assumption that requires instruments to be associated with the outcome only through exposure, we excluded SNPs strongly associated with the outcome. The effects of SNPs on exposure and outcome were then harmonized to ensure that the β values were signed to the same alleles.



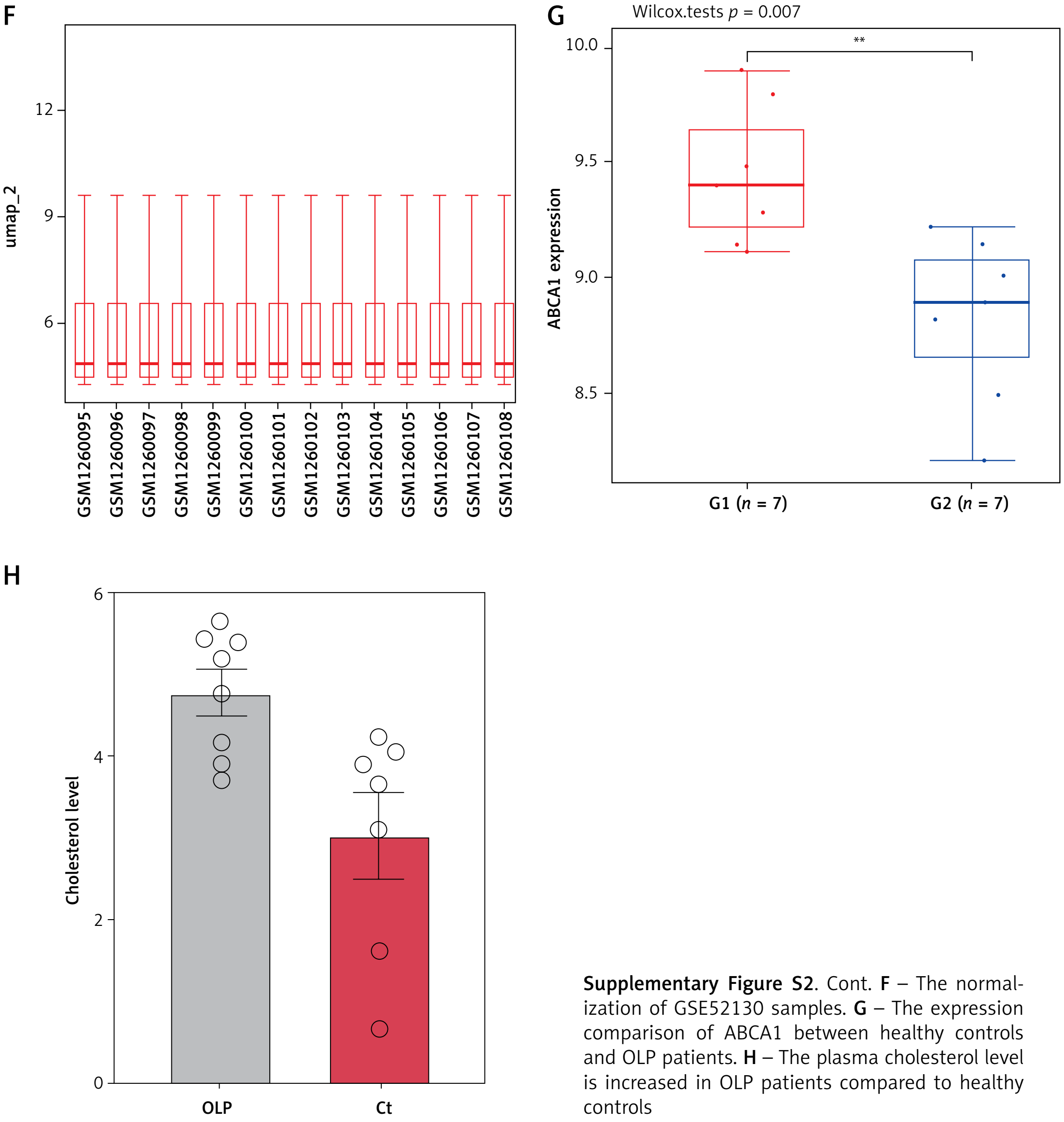
Supplementary Figure S1. The cholesterol-related genes in cell clusters. **A** – The module score in different clusters. **B** – The receptor-related gene expression among different cell clusters. **C** – The identification of cholesterol-related genes in monocytes



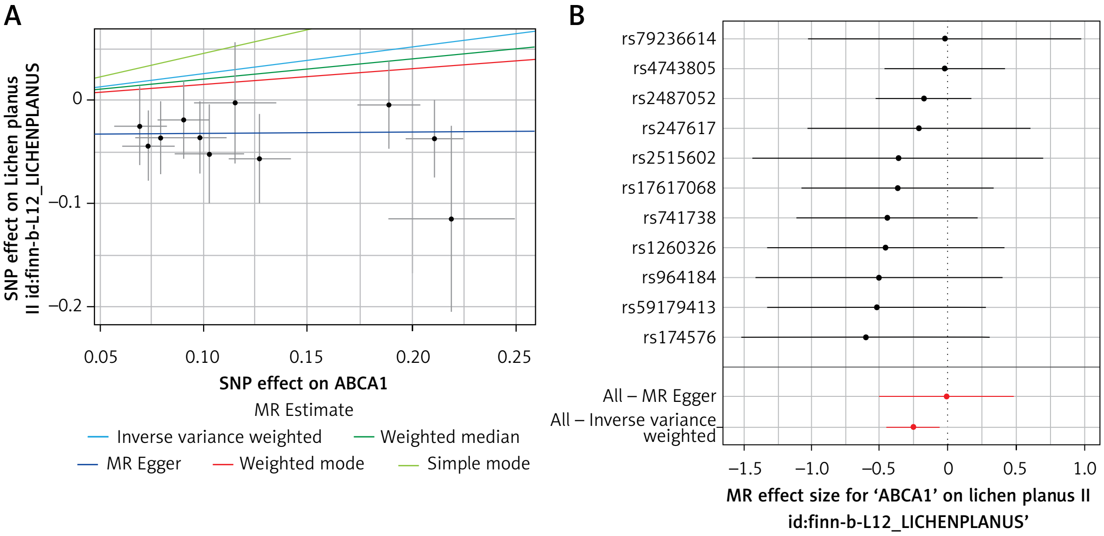
Supplementary Figure S1. Cont. **D**, **E** – The expression of ABCA1 between receptor-high monocytes and receptor-low monocytes



Supplementary Figure S2. The validation of ABCA1 in single cell analysis and bulk-RNA sequencing. **A** – The GSVA analysis for upregulated and downregulated genes in receptor-high vs. receptor-low monocytes. **B** – The UMAP of the subcluster for monocytes. **C** – The UMAP of the cholesterol-receptor gene in monocyte subclusters. **D** – The UMAP of ABCA1 expression in monocyte subclusters. **E** – The RNA velocity map in subclusters for monocytes based on ABCA1 expression



Supplementary Figure S2. Cont. **F** – The normalization of GSE52130 samples. **G** – The expression comparison of ABCA1 between healthy controls and OLP patients. **H** – The plasma cholesterol level is increased in OLP patients compared to healthy controls



Supplementary Figure S3. The MR analysis for ABCA1 eQTLs and OLP FinnGen GWAS. **A** – The scatter plot shows the positive relationship between ABCA1 eQTL and OLP. **B** – The forest plot shows the positive relationship between ABCA1 eQTL and OLP