Background:
Multiple myeloma (MM) is the second most common blood malignancy, caused by an uncontrolled growth of plasma cells in the bone marrow, accounting for 20% of all newly diagnosed hematological cancers. Although the current 5-year survival rate is ranging between 40-60%, MM is still considered an incurable disease since most of the patients eventually relapse. While the causes of MM are incompletely understood, several genome-wide association studies (GWAS) have been conducted to identify germline variants that predispose to MM. Up to date, a total of 24 loci were found to be associated with MM risk, but very little information is available about their functional role.

Aims: The principal goal is to explore in silico the function of the germline variants associated with MM risk. As we do not know if the GWAS-identified SNPs are the causal variants or just markers of risk, the functional characterization of the causal risk variants would lead to a better understanding of disease development.

Methods:
GWAS design takes advantage of the linkage disequilibrium (LD) structure of the human genome, thus the main GWAS findings are single-nucleotide polymorphisms (SNPs) that show the strongest association with MM risk (measured as the lowest p-values), but they are not necessarily the functionally causal variants. In this project, we used bioinformatics tools (GTEx, HaploReg v4.1, Roadmap, LDlinke, SNPnexus, RegulomeDB 2.0.3, SNP2TFBS, miRNASNP v3, GeneMANIA) to perform fine mapping of all GWAS-identified loci and to prioritize in each locus the polymorphism with the highest chance of being functionally relevant. In particular, we focused on the loci with the smallest number of SNPs in high LD (r^2>0.8) in order to maximize the probability to capture the casual variant.

Results:
Four of the 24 MM risk loci had a relatively small number of SNPs in high LD and within them we found that the locus located at chromosome 16 contained the greatest number of functionally annotated SNPs. Of particular functional interest was rs3747481 (chr16:30666367 C/T) due to the following reasons: it is a missense variant (protein change: P359L), it is located in the PRR14 gene, that contributes to chromatin hierarchical organization and has a role in gene regulation, has a high CADD PHRED score (22.1). Additionally, according to GTEx portal, rs3747481 is associated with the expression level of the RNF40 gene (which plays a central role in histone code and gene regulation) in whole blood cells (p=3.02^-12). It has a score of “1d” in RegulomeDB (eQTL+ TF binding + any motif + DNase peak), meaning that this variant has a high likelihood to affect binding of transcription factors. Some other SNPs (like rs35629860 and rs67128646) in the same LD block show the co-occurrence of H3K4me3 and H3K27me3 histone marks (associated with gene activation and repression, respectively) in promoters and enhancers in B lymphocytes. rs6565197 is predicted to affect the binding of the KLF4 and KLF5 transcription factors which play key roles in cell cycle regulation.

Summary/Conclusion:
Through a fine mapping of MM risk loci by bioinformatics tools, we found a variant in the locus 16p11.2 that shows in silico a very high probability to have biological role in the risk disease.