A strategy to select polymorphisms with potential impact on clinical phenotypes using open-source bioinformatic tools

S. Gagno[1;2], F. Rizzolio[1], C. Zanusso[1], S. Boffo[1], E. Dreussi[1], F. Sartor[1], G. Giorda[1], G. Toffoli[1]
[1] Experimental and Clinical Pharmacology Unit CRO-National Cancer Institute, Aviano, Italy
[2] Doctoral School in Pharmacological Science, University of Padova, Italy

Introduction: The immune system seems to be involved in the control of cancer onset and progression. Single nucleotide polymorphisms (SNPs) in immune system-related genes could be responsible for altered immune reactions and might represent a risk factor for cancer development and progression. Moreover, such SNPs might have an impact on tumor response to immune-therapies as well as conventional chemotherapy. In the rapidly-evolving pharmacogenetic field, the number of genomic information and open-source bioinformatic tools are continuously increasing, making the selection of SNPs of potential interest for tailoring therapy more complex.
In this scenario, a strategy to rationally identify a panel of germline immune system-related SNPs with potential clinical effect on ovarian cancer, has been developed. To this aim, we defined a method for SNPs-selection to design a custom kit of 192 SNPs analyzable by Illumina GoldenGate Genotyping Assay.

Methods: Immune system-related genes were selected by a literature analysis, according to the relevance of the biological process they are involved in, using 'immune system' and 'cancer' as keywords. For each gene we considered the coding sequence defined by UCSC genome browser, extending it for further 5 kbp up- and downstream to include regulatory sequences in 3'- and 5'-untranslated regions (UTRs). Using HapMap website and Haploview software, we defined blocks of SNPs in strong LD setting $r^2 \geq 0.80$ and MAF$\geq$0.05 in Caucasian population. Due to the high number of SNPs identified, we selected one SNP per block to obtain a fine mapping of every gene. To prioritize SNPs, each of them was characterized by HaploReg v2 as synonymous, missense, nonsense, residing in splice sites, introns, or UTRs, in addition to functional information concerning:

1. if the polymorphic region is an enhancer or a promoter
2. if there are transcriptional factors (TFs) or proteins bound at the polymorphic site
3. if the SNP changes a motif recognized by a transcriptional factor.

The SNPs' list was uploaded to Illumina website for the prediction of assays' designability and call-rate, each one defined by a score.
For each gene, missense SNPs and SNPs previously associated with cancer or immune system diseases were selected. Furthermore, we have applied a flowchart that prioritizes: SNPs altering a motif recognized by a TF bound in that region, polymorphic sites bound by proteins and SNPs in a promoter or enhancer sequence, according to the best final score and designability for GoldenGate Assay. In case of indecision, SNPs with the best characterization of each parameter (i.e.: the number of TFs bound and altered motifs, the strength of promoters and enhancers) were selected.

Results: Selected genes were: CD276, CXCR7, FAS, FOXO3, FOXP3, IFNG, IFNGR1, IFNGR2, IL15RA, IL17A, IL17F, IL2RA, IL2RB, IL2RG, IL8, MIF, MMP3, PRDM1, SMAD3, SMAD4, STAT3, STAT5A, STAT5B, STAT6, TGFBR1, TGFBR2, TIMP1, TIRAP, TLR10, TLR3, TLR4, TLR6, VEGFA, WNT5A. By Haploview software, we defined a list of 1170 SNPs in blocks of strong LD. 192 SNPs analyzable by GoldenGate Genotyping Assay were selected applying our flowchart. The list includes:

- 13 missense SNPs
- 16 cancer or immune diseases-associated SNPs
- 21 SNPs predicted as changing a regulatory motif
- 5 SNPs predicted as altering a promoter site
- 87 SNPs predicted as modifying an enhancer site
- 18 SNPs predicted as lying in enhancer and promoter sites at the same time
- 32 SNPs selected because of their best characterization of each parameter.

Conclusion: As genomic and functional information about polymorphisms are increasingly available on websites, their management is becoming challenging. A strategy to rationally choose the best polymorphism as a possible pharmacogenetic biomarker is required to optimize genetic analysis. Using our flowchart and open-source tools we defined a panel of 192 SNPs in candidate immune system genes that will be screened in an ovarian cancer cohort of 250 patients.