

# Melanoma expression analysis with Big Data technologies

Alicia Fernandez-Rovira<sup>1</sup>, Rocio Lavado<sup>2</sup>, Miguel Ángel Berciano Guerrero<sup>2</sup>, Ismael Navas-Delgado<sup>1</sup>, and José F. Aldana-Montes<sup>1</sup>

<sup>1</sup>Khaos Research, Universidad de Málaga, Málaga, Spain

<sup>2</sup>Unidad de Oncología Intercentros, Hospitales Univesitarios Regional y Virgen de la Victoria de Málaga, Instituto de Investigaciones Biomédicas (IBIMA), Málaga, Spain

Corresponding author:  
Ismael Navas-Delgado<sup>1</sup>

Email address: ismael@lcc.uma.es

## ABSTRACT

Melanoma is a highly immunogenic tumor. Therefore, in recent years physicians have incorporated drugs that alter the immune system into their therapeutic arsenal against this disease, revolutionizing in the treatment of patients in an advanced stage of the disease. This has led us to explore and deepen our knowledge of the immunology surrounding melanoma, in order to optimize its approach.

## INTRODUCTION

At present, immunotherapy for metastatic melanoma is based on stimulating an individual's own immune system through the use of specific monoclonal antibodies. The use of immunotherapy has meant that many of patients with melanoma have survived and therefore it constitutes a present and future treatment in this field [1 -5].

At the same time, drugs have been developed targeting specific mutations, specifically BRAF, resulting in large responses in tumor regression (set up in this clinical study to 18 months), as well as a higher percentage of long-term survivors [6-9]. These large responses trigger the massive antigens that are needed to reignite a state of immunocompetence in the patient, as well as activating biomarkers that can guide response prediction.

Unfortunately, the recent interest in developing drugs against melanoma has not been accompanied by the identification of biomarkers that are predictive of response or toxicity to them. Although preliminary data determined immunohistochemistry was useful in the tumor expression of the PDL-1 protein, subsequent work has failed to validate this predictive biomarker or prognosis. In fact, one of the main problems of the study of the immune system is its plasticity and adaptive nature. Tumor infiltrating lymphocytes (TIL), another possible postulated and non-predictive marker validated, PDL-1 expression can be varied in short time frames and this expression changes when biopsies are performed serially over time on the same tumor [10]. Therefore, more biomarkers are needed to optimize the sequences of treatments against melanoma. For these reasons, our work aims to combine both concepts: the use of therapy directed against the BRAF mutation and the reactivation of acquired immunity against melanoma.

It would therefore be useful to look for these changes in blood biomarkers in relation to immunological mechanisms, to identify predictive markers of response to new treatments, in order to identify long-term survivors with targeted therapy. Unfortunately many of these patients will not benefit from these medicines, as they are difficult to fund in terms of human and economic expenditure. However, this project seems to us to be an essential first step and the finding of these biomarkers is yet to be explored.

The analysis of the gene expression changes and their correlation with clinical changes can be developed using the tools provided by those companies which currently provide gene

expression platforms. The gene expression platform used in this clinical study is NanoString, which provides nCounter. However, nCounter has some limitations as the type of analysis is restricted to a predefined set, and the introduction of clinical features is a complex task. This paper presents an approach to collect the clinical information using a structured database and a Web user interface to introduce this information, including the results of the gene expression measurements, to go a step further than the nCounter tool.

As part of this work, we present an initial analysis of changes in the gene expression of a set of patients before and after targeted therapy. This analysis has been carried out using Big Data technologies (Apache Spark [11]) with the final goal being to scale up to large numbers of patients, even though this initial study has a limited number of enrolled patients (12 in the first analysis). This is not a Big Data problem, but the underlying study aims at targeting 20 patients per year just in Málaga, and this could be extended to be used to analyze the 3.600 patients diagnosed with melanoma per year. Besides, as the software has been developed to be applicable to any gene panel, it could be applied in any study of other diseases.

## METHODS

In this section we present the three main steps that have been developed in this project until now: clinical data collection, gene expression data collection and gene expression data analysis.

In order to collect the required data, we have developed a structured database to store clinical and genomic data. Figure 1 shows part of the database schema, where we have collected: Minimum set of personal data (gender, date of birth, medical records, medication, diagnosis date, metastasis diagnosis date, etc.); The response to the treatment; The cancer's progress, including the different visits to the hospital and the blood test values at each visit; The BRAF mutation; Primary location of the cancer; The files associated with the gene expression analysis.

The clinical information is collected by the patient's physician (previously registered and accepted in the clinical assay by the project coordinator). Thus, the effort of collecting the data is shared by all the physicians participating in the project.

The files obtained from the sequencing process are loaded by the biologist (previously registered in the system). These files are obtained from the NanoString platform. The NanoString platform can analyze 12 samples in each cartridge. As a result we obtain 12 RCC files with the gene counts for each of the gene panels. Each file is associated with the gene expression of a patient at a specified stage of treatment. These files are locally stored with a patient code, in addition to being annotated with the patient code from the project database. Then, the biologist uploads the files attached to the patient profile, indicating the date when the sample was collected and analyzed.

We have used the Immune Profiling Panel Nanostring<sup>TM</sup> (770 genes), as it has been specifically designed for cancer projects where immune aspects are studied. This panel includes 24 different immune cell types, common checkpoint inhibitors, CT antigens, and genes, covering both the adaptive and innate immune response.

Using the gene expression files, a set of analysis tools has been developed to discover patterns in the change of the gene expression levels. However, these files need to be pre-processed, as NanoString returns the flat counts of the gene expressions. The pre-processing is done to normalize the counts according to a set of control measures. The normalization process is described in [https://www.nanostring.com/download\\_file/view/251/3779](https://www.nanostring.com/download_file/view/251/3779). This process has four steps:

- Step 1. Calculate the average of the expression of each positive control in all the samples.
- Step 2. Calculate the average of the results from the previous step.
- Step 3. Calculate the relationship between specific average and global average to obtain the lane-specific scaling factor of each sample.
- Step 4. Multiply the expression of the genes by the lane-specific scaling factor.

Using the normalized gene expression values, we have produced four different representations to help clinicians analyze the expression results. These visualizations are interactive

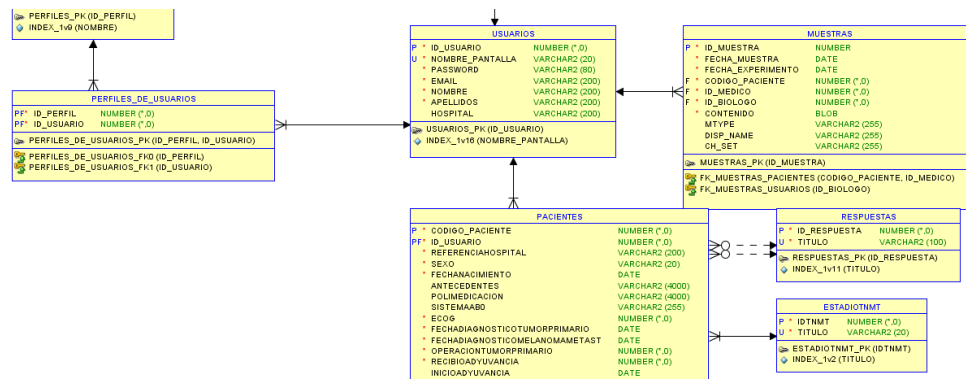


Figure 1. Extract of the database

Web graphs that can be explored by filtering the genes to be explored (just selecting a portion with the mouse). In all cases we have used only 5% of the genes with more variance in the expression:

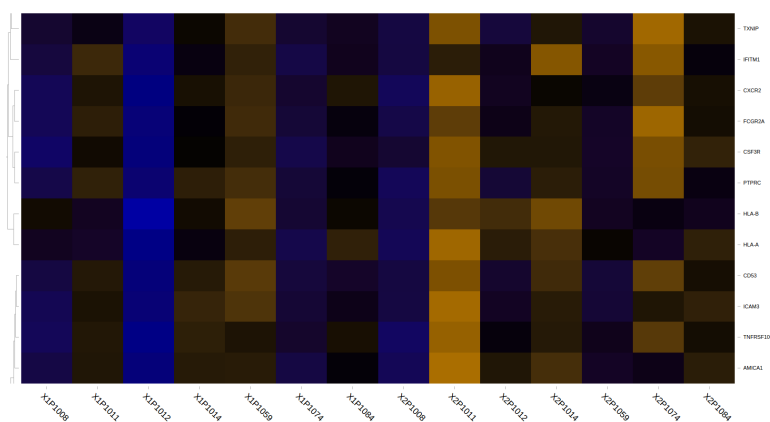
- Type I. Comparison of the samples for a single patient (patient evolution). The visualization of the expression change in a patient enables the individual analysis of the patient data with a high degree of detail.
- Type II. Comparison of the basal status of the patients (classification of the patients depending on the initial expression profile). This visualization enables the analysis of the gene expression profiles to allow users to analyze those patients with a specific pattern. For example, a user could select those patients with certain genes with a lower expression than the rest of the patients, to later explore the possible expression growth.
- Type III. Comparison of the second sample of the patients (classification of the patients according to their expression after the evolution). This visualization also allows comparisons of patients after the treatment.
- Type IV. Comparison of the patients including both samples (profiling on the patient evolution in the gene expression). This visualization type is the most complete one provided by this tool. It enables the classification and visualization of differences between patients in the evolution of the gene expression. In this case the patient samples are ordered, first showing the initial sample of all the patients, and then the results for the second sample.

## RESULTS

Using the four types, a first analysis of the patient evolution has been completed. Figure 2 shows the first results for the fourth type for the patients sequenced until now. These initial results show how genes related to the immune system have been activated in most of the patients as expected. However, in one of the cases (patient P1059) the effect is the opposite. A first qualitative analysis relates this to a previous immune activation in the patient.

The technology used has shown the possibility to scale up to larger numbers of patients, but it needs more testing. However, the development of these visualizations in a easy-to-use user interface will help clinical researchers analyze this data, and will enable semi-automatic analysis, taking clinical variables into consideration.

The analysis of another set of patients has been done, but they are not included as the visualization of the data showed that they were incorrect data, and were being studied by NanoString and the biologist working on the project to detect where the errors were produced. Thus, the use of these visualizations has demonstrated their usefulness not only in the data analysis, but also in detecting erroneous expression patterns that cannot be easily detected from the raw data.



**Figure 2.** Comparison of the gene expression changes in all the sequenced patients

## CONCLUSIONS

This paper has presented the initial results from the analysis of gene expression of the NanoString platform using Big Data technologies. These results are promising, but larger scale tests must be developed once new patients have been sequenced.

These gene expression data is being collected in a structured database, and a Web user interface is being developed to enable clinical researchers to analyze these data as soon as they are available from the NanoString platform. This integrated environment additionally includes clinical features and the evolution of the patients that are collected in the same database, so there is also work to be done in a combined analysis with the gene expression.

## ACKNOWLEDGEMENTS

This work was funded in part by Grants TIN2014-58304-R (Ministerio de Ciencia e Innovación) and P11-TIC-7529 and P12-TIC-1519 (Plan Andaluz de Investigación, Desarrollo e Innovación).

## REFERENCES

1. Hodi FS, et al. Improved survival with ipilimumab in patients with metastatic melanoma. *The New England journal of medicine*. 2010 Aug 19;363:711.
2. Robert C, et al. Ipilimumab plus dacarbazine for previously untreated metastatic melanoma. *The New England journal of medicine*. 2011 Jun 30;364:2517.
3. Topalian SL, et al. Safety, activity, and immune correlates of anti-PD-1 antibody in cancer. *The New England journal of medicine*. 2012 Jun 28;366:2443.
4. Brahmer JR, et al. Safety and activity of anti-PD-L1 antibody in patients with advanced cancer. *The New England journal of medicine*. 2012 Jun 28;366:2455.
5. Hamid O, et al. Safety and tumor responses with lambrolizumab (anti-PD-1) in melanoma. *The New England journal of medicine*. 2013 Jul 11;369:134.
6. Koya RC, et al. BRAF inhibitor vemurafenib improves the antitumor activity of adoptive cell immunotherapy. *Cancer research*. 2012 Aug 15;72:3928.
7. Hooijkaas A, et al. Selective BRAF inhibition decreases tumor-resident lymphocyte frequencies in a mouse model of human melanoma. *Oncoimmunology*. 2012 Aug 1;1:609.
8. Knight DA, et al. Host immunity contributes to the anti-melanoma activity of BRAF inhibitors. *The Journal of clinical investigation*. 2013 Mar 1;123:1371.
9. Cooper ZA, et al. Response to BRAF inhibition in melanoma is enhanced when combined with immune checkpoint blockade. *Cancer immunology research*. 2014 Jul;2:643.
10. Sanlorenzo M, Vujic I, Posch C et al. Melanoma immunotherapy. *Cancer Biol Ther*. 2014;15(6):665-74.
11. Apache Spark. <http://spark.apache.org/>