High-dimensional single cell-analysis of lymphocytic activation in melanoma

Authors

Francesca Maria Bosisio* (1)(2), Asier Antoranz* (3)(4), Maddalena Maria Bolognesi (2), Clizia Chinello (2), Jasper Wouters (1), Fulvio Magni (2), Leonidas Alexopoulos (3)(4), Giorgio Cattoretti (2), Joost van den Oord (1)

(1) Laboratory of Translational Cell and Tissue Research, KUL, Leuven, Belgium

(2) Pathology, Department of Medicine & Surgery, Università degli studi di Milano-Bicocca, Milan, Italy

(3) ProtATonce Ltd, Athens, Greece

(4) National Technical University of Athens, Greece

* These authors contributed equally to this work

Corresponding author: francescamaria.bosisio@kuleuven.be

Immune-based therapies have achieved long term survival in metastatic melanoma, but checkpoint inhibitors as monotherapy yield response rates between 15-40%[1,2,3]. Low response rates, possible severe adverse effects and high cost of these drugs have prompted the search for predictive biomarkers. Recent studies have found that durable response is associated with molecular signatures enriched in immune-related genes and heavy CD8+ T cell infiltration of the tumour at the baseline[4,5,6,7,8,9,10,11]. The lymphocytic infiltrate in melanoma is a prognostic parameter reported by the pathologist as patterns of tumour-infiltrating lymphocytes (TILs). The “brisk” pattern (diffuse or complete peripheral TILs infiltration) has a better prognosis compared to the “non-brisk” (tumour areas with TILs alternate with areas without TILs) or to the “absent” pattern (no TILs or no contact with melanoma cells)[12,13,14]. There are multiple pitfalls in the purely morphological evaluation of TILs[15], but the most important one is that morphology alone cannot determine their activation status. The meaning of the word "brisk" according to the dictionary is "active, energetic”, a definition implying a functional connotation starting from a morphological evaluation. Surprisingly, this functional connotation has never been proved. The contact between cytotoxic lymphocytes (Tcy) and melanoma does not always lead to tumour eradication but, due to immune modulation, can also result in Tcy inactivation. These “exhausted” Tcy would still be morphologically present, indistinguishable from active lymphocytes. Moreover, the morphological[16] and functional[17] side of the tumour microenvironment has been separately investigated, but integration of both types of data is still lacking.

In the latest years, several methods for single cell-analysis have been implemented in order to obtain a high-resolution landscape of the tumour microenvironment[18]. According to a recent review[19], high resolution means to characterize not only the immune infiltrate but also to define the spatial distribution of each component within it, which allows to make inferences about cell-cell interactions. Nevertheless, most of these methods rely on dissociation of the cells from fresh material, an impractical option in primary melanomas, nowadays diagnosed at early stage, with very limited material, and resulting in loss of the knowledge of the spatial distribution of each component within it. Moreover, several studies for predictive biomarkers relies on peripheral blood[20]. Though, the functional status of circulating inflammatory cells can completely change entering the tumour site[21]. Intuitively, it is the behaviour of the inflammatory cells in the surroundings of the tumour that will make the difference in terms of prognosis and response to therapy.

Here, we characterize the immune landscape at single cell-level in primary melanoma based on a panel of 39 immune markers applied on one single tissue section through a high-dimensional multiplexing method[22], an RT-PCR expression evaluation and a shotgun proteomic analysis. This approach allowed us a) to further categorize the brisk and non-brisk morphological patterns of TILs into three functional categories; b) to define the correlation between T-cell activation and spontaneous melanoma regression; c) to investigate the most important inflammatory subpopulations involved in TILs exhaustion (Figure 1).

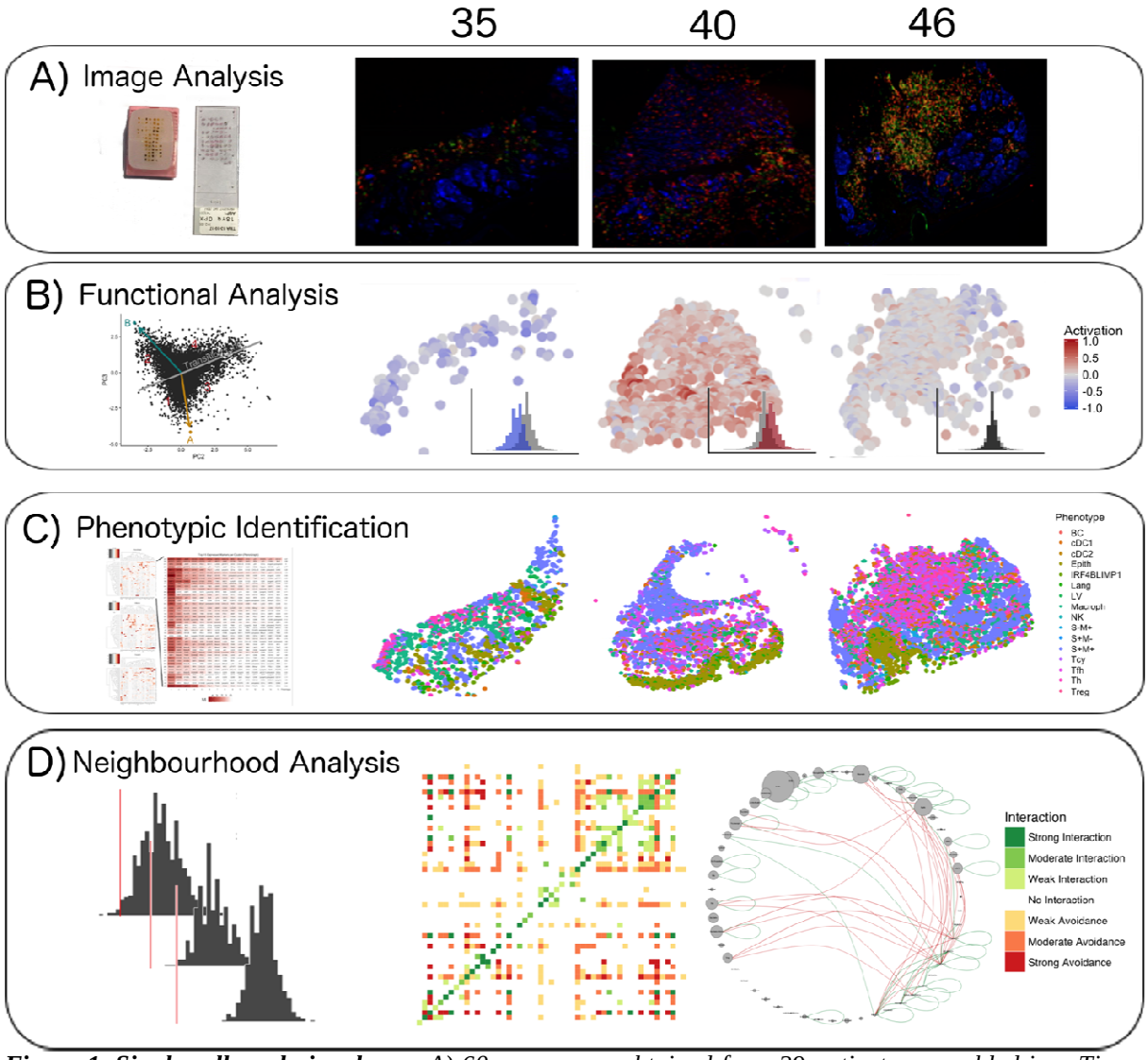

Figure 1. Single cell analysis scheme. *A) 60 cores were obtained from 29 patients, assembled in a Tissue Micro Array, and analysed using the high-dimensional multiplex method explained in Bolognesi MM, et al*[22]*; B) CD8+ cells were analysed using image analysis for functionality using an activation parameter derived from multiple activation and exhaustion markers evaluated at single cell resolution; C) All the cell populations in the cores were phenotypically identified using consensus between three clustering methods and manual annotation from the pathologists; D) The social network of the cells was analysed using a permutation test for neighbourhood analysis in order to make inferences on cell-cell interactions. All steps are exemplified using three representative images (35, 40, and 46).*

Functional analysis of TILs

We classified each CD8+ cell in the TMA cores as part of a spectrum ("cell status") ranging from "active" (CD69high and/or OX40high) to "exhausted" (TIM3highCD69lowOX40low) (Figure 2). Two intermediate states between activation and exhaustion were identified, one passing through a gradual decrease of CD69 with parallel increase of TIM3 expression, and the other in which OX40 was co-expressed with LAG3. These latter two markers had also the highest Pearson correlation coefficient among all the activation and exhaustion markers (R = 0.35). When all the cells were included in the analysis (using the DAPI mask instead of the CD8 mask), the same computational

approach repeated on the subgroup identified as Tcy (CD8+) gave a similar result, confirming the solidity of this approach (Extended Data Figure 1).

Comparing the activation level of the cells in each core to the activation level of the cells across all the cores, we assigned a label to each core in the TMA ("core status"): active (17/60 cores), in transition (23/60 cores), or exhausted (20/60 cores). At this point, with the exception of 8 patients in which the size of the melanoma allowed to sample only a single core, we could assess the heterogeneity of the immune response in different areas of the melanoma from the same patient. 10/21 showed homogeneous core statuses (4 active, 5 in transition, and 1 exhausted) and 11/21 showed heterogeneity. In summary, we observed a spatially homogeneous TILs infiltrate in 6 active cases (20.8%) and in 5 exhausted ones (17.2%), while in 18 cases (62%) the presence of areas of transition suggested an ongoing dynamic transition toward exhaustion (Figure 2.D).

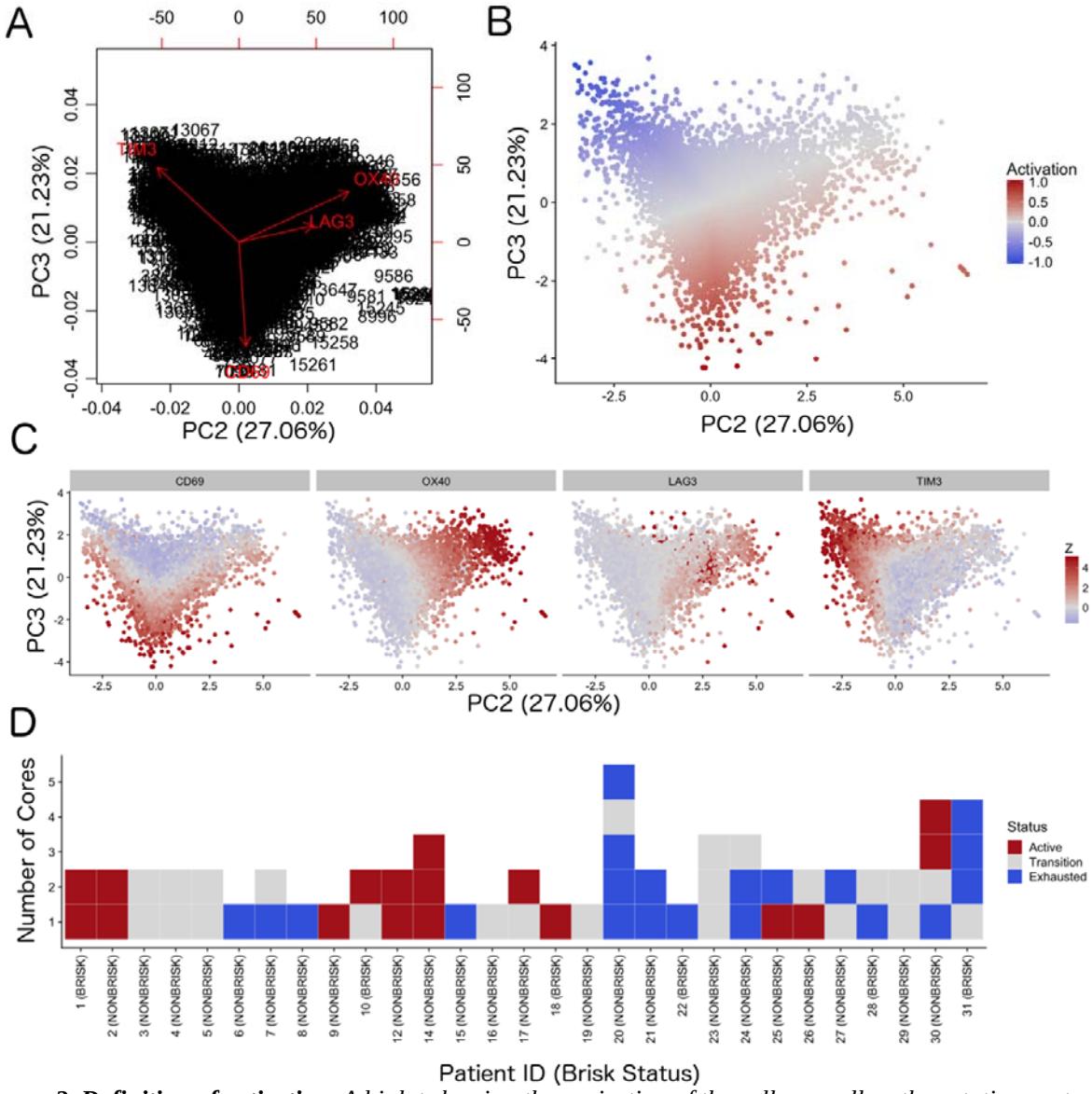

Figure 2. Definition of activation. *A biplot showing the projection of the cells as well as the rotation vectors of the markers over PC2 and PC3 (A) was the first step to define a gradient of activation going from the maximum projected value of CD69 (maximum activation) to the maximum projected value of TIM3 (maximum exhaustion) (B). C) Z-scores of the original markers over PC2 and PC3. D) Visual representation of the inter- and intra-patient heterogeneity. Each core is assigned an activation status ('Active'*

'Transition', or 'Exhausted'). The cores are grouped for each patient, giving an at-a-glance representation of the heterogeneity of the activation status in different areas of the melanoma in the same patient.

We then checked whether the cores status was significantly associated with spontaneous regression of the tumour, regarded as the result of a successful Tcy immune response, and with other histopathological parameters (histological subtype, ulceration, Breslow thickness, mitoses). Late regression areas indeed showed significant differences in the mean level of activation of the cores as compared to early regression ($p = 0.022$) and no regression ($p = 0.031$). No significant differences were instead found between early regression and no regression. Higher levels of activation were found in Lentigo Maligna Melanoma ($p = 0.02$). However, since only three LMM cores from two patients were included in our data set, no definite conclusions can be drawn from this data. The other histopathological prognostic parameters did not show significance (Extended Data Table 1).

Clustering and neighbourhood analysis

The inflammatory subpopulations were identified using three different unsupervised clustering methods (KMeans, PhenoGraph, and ClusterX) followed by manual annotation by the pathologist of the clusters and assessment of their solidity as described in the methods (Extended Data Figure 2). On 19 clusters identified, 17 could be associated to a determinate cell lineage, while the remaining two were discarded. Ultimately, we could solidly classify 179304 out of 242224 cells across all of our cores (74.02%). A further subclustering was done on the so-identified clusters using a set of functional markers with PhenoGraph, obtaining a total number of 55 subclusters.

The most abundant cell population consisted of melanoma cells (35,64%), functionally subdivided in 3 phenotypic subpopulations and 6 functional subclusters (described in Extended Data Figures 2.B and 3). We noticed that expression of MelanA in melanoma cells was associated with a more differentiated, less aggressive cell type, showing lower proliferation rates and occasional expression of p16. Moreover, these melanoma cells retained the ability to respond to IFNg, as highlighted by the expression of IFNg-related proteins and expressed less frequently immunosuppressive molecules such as PD-L1, LAG3, TIM3. The second most abundant cell type were the macrophages ("Macroph", CD68+CD163+Lysozyme+HLA-DR+, 15,24%), one quarter of them expressing at least one immunosuppressive marker: TIM3, LAG3 or PD-L1. The lymphoid compartment accounted for multiple subpopulations. Tcy (CD3+CD8+) and T helpers ("Th", CD3+CD4+FOXP3-) were the most abundant subtypes, accounting respectively for 9,48% and 8,37% of all the cells, while regulatory T cells ("Treg", CD3+CD4+FOXP3+) represented 2,35% of all lymphocytes. We interpreted the last T cell cluster as T follicular helpers ("Tfh", CXCL13+PD1+, 2,16%) even though this cluster did not express the full Tfh phenotype, (i.e. no CD3 and no CD4 expression). Within Tcy, we could identify the active (predominantly CD69+, 16,1%), transition (balanced expression of both exhaustion and expression markers, 29,6%) and exhausted (high expression of LAG3 and/or TIM3, low/absent CD69, 12,8%) functional subgroups in the Tcy cluster, plus a clonally expanding subgroup (6%), and one with low or absent expression of all functional markers, that we defined as "anergic" (34,9%). Th could also be subdivided according to their expression of activation and exhaustion markers (28,5% active Th, 18,46% Th with active/exhaustion "balanced" expression, 13,08% TIM3+ or LAG3+Th). NK cells, as expected in melanoma, were extremely infrequent (0,44%). B cells ("BC", CD20+) represented 5,03% of the total number of cells in the cores. Another cluster, CD20 negative, characterized by high expression of IRF4 and Blimp1 ("IRF4_Blimp1", 1,05%), was interpreted as most probably a subgroup of post-germinal centre BC, due to their low but still detectable expression of Bcl6 suggesting that these cells underwent a process of terminally differentiation into plasma cells[23].

Finally, we could identify among the dendritic cell group the classical dendritic cells type 1 ("cDC1", CD141+CD4+IRF8+, 3,51%), classical dendritic cells type 2 ("cDC2", CD1c+CD4+HLA-DR+, 1,99%), Langerhans cells ("Lang", CD1a+Langerin+, 1,92%), and plasmacytoid dendritic cells ("pDC", CD123+, 0,28%).

After having identified the clusters, we applied neighbourhood analysis in order to systematically identify social networks of cells to draw conclusions on actual cell-cell interactions. To understand what is the immune context that determines activation and exhaustion, we compared the results of neighbourhood analysis between the three functional subgroups (Figure 3). Most of the interactions were in common, some confirming expected findings, such as the moderate to strong interaction of Langerhans cells with the epithelium, and some revealing new information on transition from active to exhausted status, such as in the transition group Tcy interacting with TIM3+cDC1 and TIM3+macrophages, TIM3+CD69+Th interacting with TIM3+cDC1, or Tcy in general avoiding interactions with cDC2. Nevertheless, there were peculiar differences between active, exhausted and transition cases, and between brisk and non-brisk categories, that we summarized in Figure 3.B. Though, permutation tests turned out not to be the best way to investigate the interactions between melanoma cells and the surrounding cells. Melanoma cells showed avoidance from the other cells of the inflammatory infiltrate, an incoherent result due to the fact that they generally grow in nests with cells packed close to each other, there are very low chances that other interactions may result stronger than the ones in between them. Therefore, we applied another type of neighbourhood analysis as described in the methods, and we observed that macrophages and epithelial cells were in general most often located in strict proximity to the melanoma cells, without differences among the functional and morphologic categories, and brisk cases had more Tcy in close proximity with melanoma cells than non-brisk cases, as expected (Extended Data Figure 4). Interestingly, brisk cases had a higher prevalence specifically of transition and active Tcy in contact with melanoma cells compared with non-brisk cases (Active/Exhausted ratio: Brisk = 2.108762, Non-Brisk = 1.331195), that instead had relatively more exhausted and anergic cells in contact with melanoma cells (Active/Anergic ratio: Brisk = 1.743081, NonBrisk = 0.7704947).

qPCR and shotgun proteomics

To confirm the existence of the same functional subcategories in metastatic melanoma samples, that represents the type of lesions that receive the immunotherapeutic treatment we perform qPCR followed by proteomics on microdissected TILs, comparing brisk and non-brisk cases using an "absent" case and a "tumoral melanosis" (i.e. complete melanoma regression with persistence of melanin-loaded macrophages) case as controls. Furthermore, we could measure directly the levels of expression of IFNg, the best indicator of CD8+ activation, for which no suitable antibody exists for FFPE material. 4/8 melanomas with a brisk TILs pattern, and 3/7 with non-brisk TILs pattern proved active, confirming that we could subclassify from a functional point of view also metastatic lesions (Extended Data Figure 5).

To correlate the gene expression measured by qPCR with the actual protein expression we performed a proteomic analysis on the microdissected material in 3 representative cases, i.e. one with high IFNg expression ("IFNg-high"), one with the high LAG3 and no IFNg expression ("LAG3-high"), and one with none of the four markers strongly expressed ("none"). The results showed a higher number of proteins identified in the "IFNg-high" sample (324) compared to "LAG3-high" (93) and "none" (134), with almost two thirds of them (210/324) not shared with the other two samples and enriched in proteins involved in different inflammatory pathways (innate immunity, TNFR1 signalling pathway, FAS signalling pathway, T cell receptor and Fc-epsilon

receptor signalling pathway), including the interferon gamma-mediated signalling pathway. As we minimized the difference in the number of microdissected cells in each sample, these observations could only be explained by a higher production of pro-inflammatory proteins in the “IFNg-high” sample.

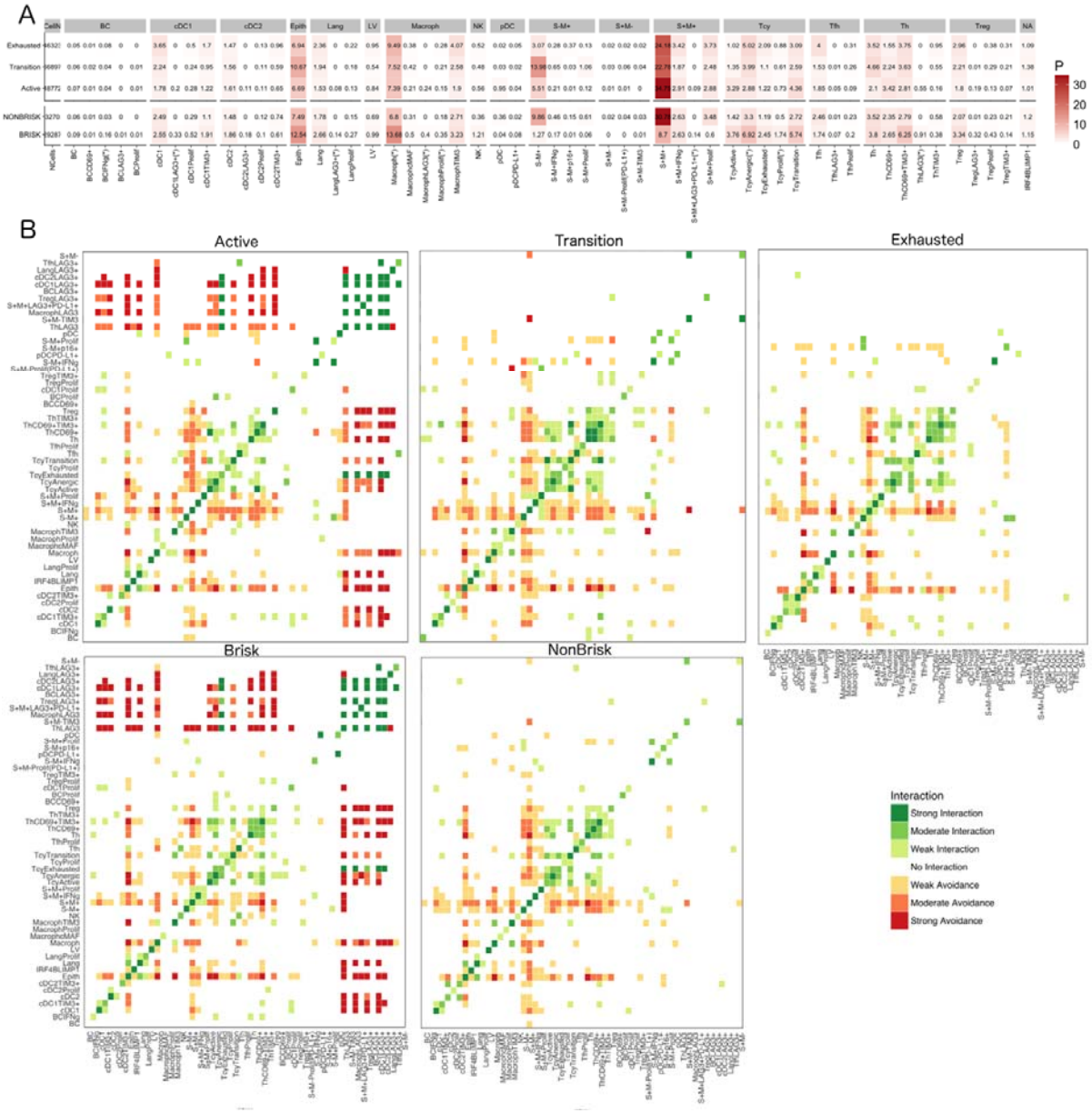

Figure 3. Phenotypic classification and neighbourhood analysis of the active, transition and exhausted microenvironment. *A) Total number of cells and the corresponding percentage for each functional group per core status (top) and brisk status (bottom). The percentages in each row adds to 100% which corresponds to the number included in the first columns (NCells); (B) The figure shows all the interactions detected with neighbourhood analysis according to the functional (top) and morphological subgroups (bottom). The interactions between different cell types are represented in a colorimetric scale (green closeness, red avoidance). In active cases the strongest interaction was observed between CD69+ Th and active Tcy; CD69+ Th were also the main subtype interacting with all the other Tcy subtypes; Treg showed no interaction with Tcy in general and very limited interactions with CD69+Th; proliferating BC were most often in contact with anergic and proliferating Tcy; macrophages showed contact with PD-L1+pDC while avoiding the other pDC; NK cells tended to avoid active Tcy and to co-localize with Tfh; Tfh contacted*

exhausted Tcy; enriched in LAG3+ cells that were very infrequent in the other two subgroups. In transition cases, Th and CD69+TIM3+Th were the main interactors with the Tcy; CD69+TIM3+Th interacted with proliferating and TIM3+Treg; BC were in contact with anergic Tcy. In both exhausted and in transition cases, CD69+TIM3+Th also interacted with IFNg BC and with TIM3+cDC2; Treg interacted with all the Th subgroups; macrophages avoided Tfh and cDC1; Tfh tended to avoid anergic/in transition T cells. In the exhausted subgroup, we had interactions between Tcy and TIM3+Th; Treg interacted with transition Tcy and anergic Tcy; there was a low-strength interaction between IFNg-BC and anergic Tcy; TIM3+ macrophages interacted with TIM3+ Treg; Tfh interacted with cDC1. The morphological subgroups "brisk" and "non-brisk" showed a mixture of interactions that could be found in active, transition and exhausted case respectively.

Discussion

The fact that not all brisk infiltrates have a good prognosis[16,24], that TILs populations can be very heterogeneous[25,26] and that anti-PD-1 immunotherapy was found to support functionally activated T cells[17] urged a thoughtful investigation of the functional status of the TILs. Since the molecules that mediate exhaustion are expressed upon activation in order to prevent the hyperactivity of the immune system[27], to investigate the activation status of lymphocytes the simultaneous expression of several molecules must be evaluated at a single cell level and at once, as a panel[28]. Moreover, the assessment of the interactions between cells requires preservation of the tissue architecture. Our study is the first to assess more than 40 markers on tissue sections at single cell resolution without losing their spatial distribution.

We obtained by high-dimensional in-situ immunotyping a snapshot of the co-expression patterns of activation and inhibition markers in tissue sections. OX40 was strongly correlated with Ki-67 and PD-1, confirming its role in sustaining Tcy clonal expansion[29]. LAG3 is expressed very early and only transiently during T cell activation, but persistent T cell activation with sustained expression of LAG3 together with other exhaustion markers (e.g. PD-L1) results in T cell dysfunction[30]. LAG3+Tregs have been shown to suppress DC maturation[31]. We found LAG3 in co-expression with PD-L1 in DC, BC, T cells, and macrophages, but not Langerhans cells, suggesting a different mechanism of immunosuppression in these cells.

We could conclude that both brisk and non-brisk cases can harbour predominantly exhausted TILs or predominantly active TILs; therefore, with the morphological classification (brisk – non brisk – absent), a complementary functional classification (active – transitional - exhausted) co-exists. The importance of going beyond the morphologic classification of TILs was previously raised by others, who claimed that a functional analysis could help in the application of an immunoscore for therapeutic prediction[32]. Only a minority (15%) of patients presented with all active cores, and 30% of patients presented with at least one active core, whereas the great majority of the patients presented with only exhausted or transition areas at the moment in which the melanoma was removed. Since the percentage of patients that obtains a durable response with single agent checkpoint therapy[1,2,3] lies between 15 and 30%, and since Krieg et al[17] reported that anti-PD-1 immunotherapy supports functionally activated T cells, it is tempting to speculate that the "mostly active" TILs cases could correspond to the durable responders to immunotherapy, while the "mostly exhausted" cases could benefit instead from combination approaches with immunotherapy and other type of therapies in order to rescue the exhausted T cells. Therefore, adding the functional evaluation could definitely improve the predictive value of the morphological TILs patterns in melanoma. Since spontaneous melanoma regression, present in only 10-35% of the melanomas, is considered to be the end-results of the melanoma-eliminating capacities of active TILs, we also studied the association of activation of TILs with early and late regression[33]. Our data showed a

clear-cut association of activation of TILs with late regression areas, indirectly proving the functional meaning of an active infiltrate.

We obtained a functional picture of the inflammatory landscape in brisk versus non-brisk cases and in active/transition/exhausted TILs microenvironments. A previous study already casted some doubts on the significance of the brisk pattern, as they found no evidence of clonally expanded TILs in some cases with a brisk infiltrate[26], and hypothesized that clonally expanded T cells might represent not only cytotoxic cells but also regulatory cells. We indeed found other T cell phenotypes increased in brisk cases, in particular NK cells, CD69+TIM3+Th, LAG3+Th, LAG3+Tfh, LAG3+Treg, and proliferating Treg. Moreover, brisk infiltrates were enriched not only with immune stimulating cells, but also with cells with an immunosuppressive role (Figure 2.A). Active and exhausted cases were instead functionally coherent categories. An active microenvironment was enriched for active Tcy, active CD69+Th and LAG3+Treg. On the other hand, an exhausted microenvironment was enriched for a wide range of immunosuppressive cells (Figure 2.A). The functional coherency of the functional classification was highlighted also by the neighbourhood analysis. While brisk and non-brisk cases both showed a mixture of activating, transitional and exhausting interactions, confirming once more that the functional and morphological categories do not bring the same type of information, but nicely complement each other, active, transitional and exhausted groups showed instead interesting and peculiar differences. From the neighbourhood analysis we could detect in exhausted and transition cases but not in active cases the specific cell types that contributed to Tcy exhaustion: BC, TIM3+cDC2, and Treg (Figure 2.B). In active cases, the interactions between the active T-cell subtypes were stronger than in transition and exhausted cases, where interactions between Th and Tcy were characterized by expression of the inhibitory marker TIM3. Interestingly, LAG3+ cDC1, cDC2, Treg, macrophages and Th were arranged in small clusters that were found only in active and brisk cases, where they strongly interacted with exhausted T cells and avoided interactions with other cells. These LAG3+ clusters may, therefore, be an important factor involved in the early onset of the exhaustion program, and they may represent one of the most interesting targets to prevent the formation of an exhausted microenvironment.

In conclusion, in this paper we have shown that the activation status of the TILs does not necessarily parallel the morphological categories, and that within a single melanoma, the inflammatory response may vary considerably. We hypothesize that the general good prognosis of melanomas with a brisk pattern of TILs could be based on the fact that the melanoma is excised and analysed at a moment in which it is still in control by the inflammatory cells that are still abundant in the immune microenvironment. Therefore, quantity, in addition to quality, may be an important determinant for the prognosis (as previously stated[34]) who added a quantitative dimension to the morphological categories), but not for the prediction of the response to therapy. This should be taken in consideration when assessing the response to immunotherapy and will ultimately lead to the identification of functional inflammatory microenvironments that may benefit from personalized combined therapy protocols.

METHODS

Sample Description

Twenty-nine invasive primary cutaneous melanomas from the Department of Pathology of the University Hospitals Leuven (KU Leuven), Belgium, were classified based on the H&E staining according to the pattern of the inflammatory infiltrate into brisk (6 cases) and non-brisk (23 cases). According to their subtype, 24 superficial spreading melanomas, 3 nodular melanomas, and 2 lentigo maligna melanoma were included.

TMA construction

Tissue Micro Arrays (TMAs) were constructed with the GALILEO CK4500 (Isenet Srl, Milan, Italy). For each patient, one to five representative regions of interest were sampled according to the size of the specimen and the morphological heterogeneity both of the melanoma and of the distribution of the infiltrate. According to the morphological TILs pattern, we sampled brisk areas (i.e. 6 brisk areas in melanomas with brisk TILs pattern, termed “brisk in brisk” and 10 brisk areas in melanomas with non-brisk TILs pattern, termed “brisk in non-brisk”) and non-brisk areas (i.e. 15 non-brisk areas in non-brisk melanomas, termed “non-brisk in non-brisk”); in addition, areas showing “early regression” (7), ” late regression” (5) and “no regression” (17) according to current morphological criteria[33] were sampled. After processing, cutting and staining of the TMA blocks, a total of 60 cores were available for analysis.

Multiplex-Stripping Immunofluorescence

The multiplex-stripping method was performed as published in Bolognesi MM, et al[22]. The antibodies used for the multiplex staining are listed in the Extended Data Table 2. Slides were scanned, and images acquired with the Hamamatsu Nanozoomer S60 scanner (Nikon, Italia) equipped with a Nikon 20X/0.75 lambda objective. Complete removal of previous layers was monitored by: a) checking for unexpected staining consistent with the subcellular and tissue distribution of a previously detected marker; and b) spurious co-clustering of unrelated molecules in the hierarchical clustering images (see below).

Image Pre-processing

Fiji/ImageJ (version 1.51u) were used to pre-process the images (File format: from. ndpi to.tif 8/16 bit, grayscale). Registration was done through the Turboreg and MultiStackReg plugins, by aligning the DAPI channels of different rounds of staining, saving the coordinates of the registration as Landmarks and applying the landmarks of the transformation to the other channels. Registration was followed by autofluorescence subtraction (Image process → subtract), previously acquired in a dedicated channel, for FITC, TRITc and Pacific Orange. A macro was written in Fiji/ImageJ and used for the TMA segmentation into single images. Cell segmentation, mask creation, and single cell measurements were done with a custom pipeline using CellProfiler (version 2014-07-23T17:45:00 6c2d896). Quality Control (QC) over the Mean Fluorescence Intensity (MFI) values was performed using feature and sample selection. In short, those cells that did not have expression in at least three markers, and those markers that were not expressed in at least 1% of the samples were removed. MFIs were further normalized to Z-scores as recommended in Caicedo JC, et al[35]. Z-scores were trimmed between -5 and +5 to avoid a strong influence of any possible outliers in the downstream analysis. The correlation between the different markers was calculated using Pearson’s correlation coefficient.

Functional analysis of TILs

We selected two activation (CD69, OX40) and two exhaustion (TIM3, LAG3) markers after literature review and preliminary testing of the antibody performance on control FFPE under the conditions of the multiplex protocol. The expression levels of these markers were measured selectively on CD8+ lymphocytes using a first mask focused only on CD8+ cells. Principal Component Analysis (PCA) was applied over the expression values to evaluate the functional structure of the data and to assign an activation value in the [-1, 1] range to each cell (Supplementary Information 1, Extended Data Figure 6). Briefly, principal components (PCs) 2 and 3 were used as the rotation matrix revealed that PC1 contained all the markers in the same direction (same sign). The point of maximum activation (Activation = 1), was defined where the projected value of CD69 (marker of activation) over PCs 2 and 3 was at the maximum while the point of maximum exhaustion (Activation = -1) where the projected value of TIM3 (marker of exhaustion) over PCs 2 and 3 was at the maximum (Figure 2). The gradient of transition was defined between the previously defined points and the centroid of the projected dataset. Pairwise t-tests with pooled standard deviation (sd) were used to find significant differences in the level of activation of the images regarding multiple histopathologic parameters (brisk/non-brisk infiltrate, regression, number of lymphocytes, ulceration, Breslow thickness, mitoses, subtype of melanoma). For this purpose, the activation level of each image was represented by the mean of its cells while the intrinsic degree of heterogeneity was captured by the sd. Continuous values of the histopathological parameters were fitted using linear models instead. P-values were adjusted for multiple comparisons using the holm method. A cut-off of 0.05 was used as significance threshold for the adjusted p-values. Images were further classified into: ‘Active’, ‘Transition’, and ‘Exhausted’ (from now on, indicating the core status) using one-tailed t-tests comparing the distribution of the activation values in a specific image versus the background distribution (combination of all images). P-values were adjusted using the False Discovery Rate (fdr) method. A cut-off value of 0.001 over the adjusted p-values was used as classification threshold.

Identification of cell subpopulations

To evaluate the cell subpopulations, a second mask based on the DAPI nuclear staining contour expanded by 5 pixels was created. A two-tier approach was followed for the identification of cell subpopulations: phenotypic and functional. The phenotypic identification was conducted by applying three different clustering methods: PhenoGraph, ClusterX, and K-means, over the phenotypic markers: CD3, CD20, CD4, HLA-DR, Bcl6, CD16, CD68, CD56, CD141, CD1a, CD1c, Blimp1, Langerin, Lysozyme, Podoplanin, FOXP3, S100AB, IRF4, IRF8, CD1a, CXCL13, CD8, CD138, CD123, PD-1, and MelanA. PhenoGraph and ClusterX were implemented using the cytofkit package from R[36]. Clusters were represented by a vector containing the mean of each marker and were used to further associate them to a cell subpopulation using prior knowledge. For a phenotype to be assigned to a cell, at least two clustering methods should agree on their predicted phenotype. Prior to functional identification, PCA was repeated over the Tcy cells (CD8+) using CD69, OX40, LAG3, and TIM3 markers in order to confirm that with the new mask, the same dataset with the same structure as with the CD8+ mask could be retrieved (Extended Data Figure 1). The functional identification was conducted by applying PhenoGraph over the functional markers: CD69, Ki-67, TAP2, GBP1, MYC, p16, MX1, OX40, c-Maf, PD-L1, LAG3, TIM3, and Phospho-Stat1 (with the exception of Tcy cells for which we used a personalized panel consisting of: CD8, CD69, OX40, LAG3, TIM3, PD-1, and Ki-67). Clusters were represented, associated to cell subpopulations, and evaluated for stability as described for the phenotypic identification. One-way Analysis of Variance (ANOVA) models were fitted to identify if the percentage of cells in any

subpopulation was linked to the core status. The same approach was repeated for the brisk infiltrate histopathologic parameter.

An unbiased quantitative analysis of cell-cell interactions was performed using an adaptation of the algorithm described in Schapiro D, et al[37] for neighbourhood analysis to systematically identify social networks of cells and to better understand the tissue microenvironment. Our adaptation also uses a kernel-based approach (radius = 30 px) to define the neighbourhood of a cell and a permutation test (N = 1000) to compare the number of neighbouring cells of each phenotype in a given image to the randomized case. This allows the assignment of a significance value to a cell-cell interaction representative of the spatial organization of the cells. Significance values were further classified into: avoidance (-1), non-significant (0), and proximity (1) using a significance threshold of 0.001 (more significant that all the random cases). Interactions across images were integrated according to equation 1:

$$P_{i,j} = \frac{\sum_{k=1}^{M}(c_{i,j,k} \cdot \sqrt{N_{i,j,k}})}{\sum_{k=1}^{M}(\sqrt{N_{i,j,k}})}$$

where *Ci,j,k* is the significance value (-1, 0, or 1) of the interaction between cell types *i* and *j* for image *k*, and *Ni,j,k* is the geometric average of the number of cells of type *i* and *j* for image *k*. Cell-cell interactions were considered strong if they were significant in at least 75% of the N-adjusted cases (abs(P) > 0.75), moderate if 50%, (0.5 < abs(P) <= 0.75), weak if 25% (0.25 < abs(P) <= 0.5), and non-significant otherwise (abs(P) <= 0.25). A comparative analysis of the above described method was performed for the different core statuses as well as for the different brisk infiltrate cases (Figure 3.B).

Even though neighbourhood analysis allows evaluation of cell-cell interactions, the mathematical model applied is limited in cases where there are dominant cell types that grow in nests, with cells packed next to each other. Therefore, for melanoma cells, we evaluated the closest neighbours by counting the number of cells of each subpopulation (apart from melanoma cells) that were in their neighbourhood and divided the amount by the geometric average of the number of melanoma cells and the number of cells of the specific population across all the cores in which the specific population appears. This analysis was repeated for the different brisk and activation cases.

Laser Microdissection

18 fresh frozen melanoma metastases with different types of TILs patterns (9 brisk, 7 non brisk, 1 absent and 1 tumoral melanosis) were collected in the Department of Pathology of the University Hospitals Leuven (KU Leuven), Belgium. 10 micrometre-thick sections were cut from each fresh frozen block and put on a film slide (Zeiss, Oberkochen, Germany). Sections were stained with crystal violet. Areas with dense TILs infiltrate were microdissected with the Leica DM6000 B laser microdissection device (Leica, Wetzlar, Germany). A calculation was made in order to microdissect the same surface in all the samples in order to minimize differences between the samples (around 10.000 lymphocytes/sample).

qPCR of laser microdissected samples

RNA extraction was done with RNeasy® Plus Micro Kit (Qiagen) according to the protocol. cDNA retrotranscription followed by an amplification step was done with Ovation® Pico SL WTA System V2 (Nugen) according to protocol. Primers for Interferon gamma (INFg, forward ‘TGTTACTGCCAGGACCCA’ and reverse ‘TTCTGTCACTCTCCTCTTTCCA’), TIM3 (forward ‘CTACTACTTACAAGGTCCTCAGAA’ and reverse ‘TCCTGAGCACCACGTTG’), LAG3

(forward ‘CACCTCCTGCTGTTTCTCA’ and reverse ‘TTGGTCGCCACTGTCTTC’), CD40-L (forward ‘GAAGGTTGGACAAGATAGAAGATG’ and reverse ‘GGATAAGGATCTTTCTCCTGTGTT’), CD45 (forward ‘GCTACTGGAAACCTGAAGTGA’ and reverse ‘CACAGATTTCCTGGTCTCCAT’), Beta2microglobulin (forward ‘ACAGCCCAAGATAGTTAAGTG’ and reverse ‘ATCTTCAAACCTCCATGATGC’), HPRT (forward ‘ATAAGCCAGACTTTGTTGGA’ and reverse ‘CTCAACTTGAACTCTCATCTTAGG’) were designed with Perl Primer® and tested in our laboratory. 96-wells plates were loaded with Fast SYBR® Green Master Mix, the primers and the samples in the recommended proportions, and analysed with the 7900 HT Fast Real-Time PCR system (Applied Biosystems). The log fold change (logFC) of the expression values towards the expression value of CD45 were calculated. If the log(IFNg/CD45) was positive, the sample was classified as positive. On the other hand, exhaustion was defined by expression of LAG3 and/or TIM3 with lack of IFNg and CD40L expression.

Shotgun Proteomics of laser microdissected samples

The materials used for the shotgun proteomics analysis were: Trifluoroacetic acid, MS grade porcine trypsin, DTT (dithiothreitol), IAA (Iodoacetamide), ABC (Ammonium Bicarbonate), HPLC grade water, acetonitrile (ACN), were from Sigma-Aldrich (Sigma-Aldrich Chemie GmbH, Buchs, Switzerland). All solutions for Mass Spectrometry (MS) analysis were prepared using HPLC-grade. LCM collected material corresponding to about 10^4 cells for each sample group was re-suspended in 90 µl of bidistilled water and immediately stored at -80°C. For the bottom-up MS analysis, all the samples were processed and trypsinized. Briefly, thawed cells were submitted to a second lysis adding 60 µl of 0.25% w/v RapiGest surfactant (RG, Waters Corporation) in 125mM ammonium bicarbonate (ABC) and sonicated for 10 min. Samples were then centrifuged at 14 000 × g for 10 min. About 140 µl of supernatants were collected, transferred in a new tube and quantified using bicinchoninic acid assay (Pierce -Thermo Fisher Scientific). After 5min denaturation (95°C), proteins were reduced with 50mM DTT in 50mM ABC at room temperature and alkylated with 100mM IAA in 50mM ABC (30 min incubation in dark). Digestion of samples was performed overnight at 37°C using 2µg of MS grade trypsin. RG surfactant were removed using an acid precipitation with TFA at a final concentration of 0.5% v/v. Samples were then spun down for 10 min at 14000 x g and supernatants collected for MS analysis. Peptide mixtures were desalted and concentrated using Ziptip™ µ-C8 pipette tips (Millipore Corp, Bedford, MA). An equal volume of eluted digests was injected at least three times for each sample into Ultimate™ 3000 RSLCnano (ThermoScientific, Sunnyvale, CA) coupled online with Impact HD™ UHR-QqToF (Bruker Daltonics, Germany). In details, samples were concentrated onto a pre-column (Dionex, Acclaim PepMap 100 C18 cartridge, 300 µm) and then separated at 40°C with a flow rate of 300 nL/min through a 50 cm nano-column (Dionex, ID 0.075mm, Acclaim PepMap100, C18). A multi-step gradient of 4 hours ranging from 4 to 66% of 0.1% formic acid in 80% ACN in 200min was applied[38]. NanoBoosterCaptiveSpray™ ESI source (Bruker Daltonics) was directly connected to column out. Mass spectrometer was operated in data-dependent acquisition mode, using CID fragmentation assisted by N2 as collision gas setting acquisition parameters as already reported[39]. Mass accuracy was assessed using a specific lock mass (1221.9906 m/z) and a calibration segment (10 mM sodium formate cluster solution) for each single run. Raw data from nLC ESI-MS/MS were elaborated through DataAnalysis™ v.4.1 Sp4 (Bruker Daltonics, Germany) and converted into peaklists. Resulting files were interrogated for protein identification through in-house Mascot search engine (version: 2.4.1), as described[38]. Identity was accepted for proteins recognized by at least one unique and significant (p-value < 0.05) peptide.

Pathways Analysis

Gene-set enrichment analysis was performed with DAVID 6.8[40,41]. Pathways were visualized and partially analysed with STRING v10[42].

Acknowledgements

Francesca Maria Bosisio is funded by the MEL-PLEX research training programme ('Exploiting MELanoma disease comPLEXity to address European research training needs in translational cancer systems biology and cancer systems medicine', Grant agreement no: 642295, MSCA-ITN-2014-ETN, Project Horizon 2020, in the framework of the MARIE SKŁODOWSKA-CURIE ACTIONS).

Asier Antoranz is funded by the SyMBioSys research training programme ('Systematic Modeling of Biological Systems'), grant agreement no: 675585, MSCA-ITN-2015-ETN, Project Horizon 2020, in the framework of the MARIE SKŁODOWSKA-CURIE ACTIONS.

Maddalena Maria Bolognesi is supported by a clinical research project BEL114054 (HGS1006-C1121) of the University of Milano-Bicocca and GlaxoSmithKline, UK.

The authors would also like to thank Theodore Sakellaropoulos for his feedback in designing the computational pipeline.

Author Contributions

F.M.B. designed the study, executed qPCR and microdissection, did pathways analysis, contributed to design the pipeline for pre-processing of the raw data (images) and processed all the images, participated in directing the data analysis pipeline, manually annotated all the clusters, coordinated the different expertise in the group and wrote the paper.

A.A. did all the programming in R and created the bioinformatics pipeline for data analysis choosing the most appropriated analytical methods (except for pathway analysis), analysed the data from qPCR and wrote the paper.

M.M.B. designed the pipeline for pre-processing the raw data in order to obtain the data for data analysis and reviewed the paper.

C.C. executed the shotgun proteomics and preliminary analysed the raw data up to pathways analysis, wrote the part on shotgun proteomics and reviewed the paper.

J.W. gave essential help with the qPCR and the microdissection and reviewed the paper.

F.M. gave important advice on the planning of the proteomics experiments and reviewed the paper.

L.A. supervised the computational pipeline and provided insightful feedback during the writing of the paper.

G.C. created and optimized the MILAN multiplex staining method (Cattoretti G et al., Multiple Iterative Labeling by Antibody Neodeposition (MILAN), Protocol Exchange (2018) doi:10.1038/protex.2018.106), practically executed all the stainings and scanned the slides generating the raw data, gave important input on the perspective of the paper and reviewed the paper.

J.v.d.O. suggested the topic, participated in designing the study, gave essential input and direction for the study and supervised the experiments, reviewed the paper.

References

1. Ribas, A. & Flaherty, K.T. Gauging the long-term benefits of Ipilimumab in melanoma. *J. Clin. Oncol.* **33**, 1865-6 (2015).
2. Ribas, A. *et al.* Association of Pembrolizumab with tumor response and survival among patients with advanced melanoma. *JAMA.* **315**, 1600-9 (2016).
3. Robert, C. *et al.* Nivolumab in previously untreated melanoma without BRAF mutation. *N. Engl. J. Med.* **372**, 320-30 (2015).
4. Seremet, T. *et al.* Molecular and epigenetic features of melanomas and tumor immune microenvironment linked to durable remission to ipilimumab-based immunotherapy in metastatic patients. *J. Transl. Med.* **14**, 232 (2016).
5. Hugo, W. *et al.* Genomic and transcriptomic features of response to anti-PD-1 therapy in metastatic melanoma. *Cell* **165**, 35-44 (2016).
6. Gibney, G.T., Weiner, L.M. & Atkins, M.B. Predictive biomarkers for checkpoint inhibitor-based immunotherapy. *Lancet Oncol.* **17**, 542-551 (2016).
7. Teixido, C., Gonzalez-Cao, M., Karachaliou, N. & Rosell, R. Predictive factors for immunotherapy in melanoma. *Ann. Transl. Med.* **3**, 208-218 (2015).
8. Vasaturo, A. *et al.* T cell landscape in a primary melanoma predicts the survival of patients with metastatic disease after their treatment with dendritic cell vaccines. *Cancer Res.* **76**, 3496-506 (2016).
9. Loo, K. & Daud, A. Emerging biomarkers as predictors to anti-PD1/PD-L1 therapies in advanced melanoma. *Immunotherapy* **8**, 775-85 (2016).
10. Koguchi, Y. *et al.* Serum immunoregulatory proteins as predictors of overall survival of metastatic melanoma patients treated with Ipilimumab. *Cancer Res.* **75**, 5084-92 (2015).
11. Teng, M.W., Ngiow, S.F., Ribas, A. & Smyth, M.J. Classifying cancers based on T-cell infiltration and PD-L1. *Cancer Res.* **75**, 2139-45 (2015).
12. Clark, W.H. *et al.* Model predicting survival in stage I melanoma based on tumor progression. *J. Natl. Cancer. Inst.* 81, 1893-904 (1989).
13. Clemente, C.G., *et al.* Prognostic value of tumor infiltrating lymphocytes in the vertical growth phase of primary cutaneous melanoma. *Cancer* **77** 1303-10 (1996).
14. Mihm, M.C. Jr., Clemente, C.G. & Cascinelli, N. Tumor infiltrating lymphocytes in lymph node melanoma metastases: a histopathologic prognostic indicator and an expression of local immune response. *Lab. Invest.* **74**, 43-7 (1996).
15. Bosisio, F.M. & van den Oord, J.J. Immunoplasticity in cutaneous melanoma: beyond pure morphology. *Virchows Arch.* **470**, 357-369 (2017).
16. Saltz, J. *et al.* Spatial organization and molecular correlation of tumor-infiltrating lymphocytes using deep learning on pathology images. *Cell Rep.* **23**, 181-7 (2018).
17. Krieg, C. *et al.* High-dimensional single-cell analysis predicts response to anti-PD-1 immunotherapy. *Nat. Med.* **24**, 144-153 (2018).
18. Biology of single cells shines a light on collaboration. *Nature* **547**, 5 (2017).
19. Binnewies, M. *et al.* Understanding the tumor immune microenvironment (TIME) for effective therapy. *Nat. Med.* **24**, 541-550 (2018).
20. Ascierto, P.A. *et al.* Future perspectives in melanoma research. *J. Transl. Med.* **15**, 236 (2017).

21. Buggert, M. *et al.* Identification and characterization of HIV-specific resident memory CD8+ T cells in human lymphoid tissue. *Sci. Immunol.* **3**, eaar4526 (2018).
22. Bolognesi, M.M. *et al.* Multiplex Staining by Sequential Immunostaining and Antibody Removal on Routine Tissue Sections. *J. Histochem. Cytochem.* **65**, 431-444 (2017).
23. Alinikula, J., Nera, K.P., Junttila, S. & Lassila, O. Alternate pathways for Bcl6-mediated regulation of B cell to plasma cell differentiation. *Eur. J. Immunol.* **41**, 2404-13 (2011).
24. Thorsson, V. *et al.* The immune landscape of cancer. *Immunity*. **48**, 812-830 (2018).
25. Simoni, Y. *et al.* Bystander CD8+ T cells are abundant and phenotypically distinct in human tumour infiltrates. *Nature*. Epub ahead of print.
26. Bernsen, M.R. *et al*. Presence and localization of T-cell subsets in relation to melanocyte differentiation antigen expression and tumour regression as assessed by immunohistochemistry and molecular analysis of microdissected T cells. *J. Pathol.* **202**, 70-9 (2004).
27. Pao, W. *et al*. Tissue-specific immunoregulation: a call for better understanding of the "immunostat" in the context of cancer. *Cancer Discov.* **8**, 395-402 (2018).
28. Wherry, E.J. T cell exhaustion. *Nat. Immunol*. **12**, 492-9 (2011).
29. Weixler, B. *et al*., OX40 expression enhances the prognostic significance of CD8 positive lymphocyte infiltration in colorectal cancer. *Oncotarget.* **6**, 37588-99 (2015).
30. Ruea-Yea, H. *et al*. LAG3 and PD1 co-inhibitory molecules collaborate to limit CD8+ T cell signaling and dampen antitumor immunity in a murine ovarian cancer model. *Oncotarget.* **6**, 27359–27377 (2015).
31. Andrews, L.P., Marciscano, A.E., Drake, C.G. & Vignali, D.A. LAG3 (CD223) as a cancer immunotherapy target. *Immunol. Rev.* **276**, 80-96 (2017).
32. Stefani, S. *et al.* Density of immunogenic antigens does not explain the presence or absence of the T-cell–inflamed tumor microenvironment in melanoma. *Proc. Natl. Acad. Sci. USA.* **113**, E7759-E7768 (2016).
33. Botella-Estrada, R., Traves, V., Requena, C., Guillen-Barona, C. & Nagore, E. Correlation of histologic regression in primary melanoma with sentinel node status. *JAMA Dermatol.* **150**, 828-35 (2014).
34. Azimi, F. *et al.* Tumor-infiltrating lymphocyte grade is an independent predictor of sentinel lymph node status and survival in patients with cutaneous melanoma. *J. Clin. Oncol.* **30**, 2678-83 (2012).
35. Caicedo, J.C. *et al.* Data-analysis strategies for image-based cell profiling. *Nat. Methods* **31**, 849-863 (2017).
36. Chen, H. *et al.* Cytofkit: A Bioconductor Package for an Integrated Mass Cytometry Data Analysis Pipeline. *PLoS Comput. Biol.* **12**, e1005112 (2016).
37. Saphiro, D. *et al.* miCAT: A toolbox for analysis of cell phenotypes and interactions in multiplex image cytometry data. *Nat. Methods* **14**, 873-876 (2017).
38. Chinello, C. *et al.* Proteomics of liquid biopsies: Depicting RCC infiltration into the renal vein by MS analysis of urine and plasma. *J Proteomics*. Apr 22. (2018).
39. Liu, X. *et al.* Intraluminal proteome and peptidome of human urinary extracellular vesicles. *Proteomics Clin. Appl.* **9**, 568-73 (2015).
40. Huang, D.W., Sherman, B.T. & Lempicki, R.A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44 (2008).
41. Glynn, D. et al. DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol.* **4**, P3 (2003).
42. Szklarczyk, D. *et al.* STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acid Res.* **43**, D447-52 (2015).