TerraFlow, A New High Parameter Data Analysis Tool, Reveals Systemic T-cell Exhaustion and Dysfunctional Cytokine Production in Classical Hodgkin Lymphoma

Catherine Diefenbach[1]*, Daniel Freeman[1]*, Linda Lam[1], Tri Le[2], Jason Alexandre[1], Bruce Raphael[1], Michael Grossbard[1], David Kaminetzky[1], Jia Ruan[3], and Pratip K. Chattopadhyay[4]

[1] Perlmutter Cancer Center at NYU Langone Health

[2] BD Biosciences

[3]Division of Hematology and Medical Oncology, Weill Cornell Medicine and New York Presbyterian Hospital

[4]Talon Biomarkers

*Contributed equally

Correspondence: pratip@talonbiomarkers.com

ABSTRACT

The incredible variety of immune-related proteins presents enormous challenges in immuno-monitoring. Combinatorial expression of these proteins defines cell types that may influence disease. Using high-parameter flow cytometry, and a new data analysis algorithm (TerraFlow), immunophenotypes can be comprehensively surveyed for disease associations. In classical Hodgkin lymphoma, where systemic T-cell immunity has not been investigated in detail, we reveal immune perturbations in newly-diagnosed patients (compared to healthy controls): 1) reduced levels of early (CD127+ CCR7+) memory T-cells, 2) elevated levels of activated (CD278+) memory T-cells primed for apoptosis (CD95+) and expressing inhibitory/exhaustion receptors (CD272+, PD1+, CD152+, CD366+), 3) increased suppressive (GITR+) cells, and 4) a shift away from TH1 and TH2 cells (IFNg+, IL4+) toward IL17-producing cells. Many of these perturbations remain after treatment. Our results provide mechanistic support for past reports of immune deficiency in Hodgkin lymphoma, detail new immunotherapy and biomarker research targets, and suggest strategies for combination therapies.

INTRODUCTION

Immune responses are coordinated by a myriad of proteins, distributed across a wide variety of cell types. The representation of particular cell types may significantly influence the immune response in malignancy. In lymphoma, for example, the immune landscape of the tumor microenvironment plays a clear role in disease. Patients with classical Hodgkin lymphoma (cHL) possess rare, malignant Hodgkin/Reed Sternberg (HRS) cells that shape their microenvironment to prevent immune surveillance and inhibit cytotoxic immune responses[1-5]. Various mechanisms underlying this phenomena have been proposed including the secretion of inhibitory cytokines such as TARC (CCL17), attraction of suppressive T helper 2 (Th2) and regulatory T (Treg) cells to the tumor microenvironment (TME), and differentiation of naïve CD4+ T cells into forkhead box P3 (FoxP3+) regulatory (Treg) T cells[6-8].

Less, however, is known about the systemic immune system of patients with cHL. Specifically, are immune perturbations in cHL global or local, and what influence does TME exert on systemic immunity? A number of studies suggest systemic immune dysregulation in early and advanced stage cHL, as demonstrated by poor responses to recall antigens in delayed-type hypersensitivity testing[9] and systemic elevations of a variety of immune modulators, including CCL17 (TARC)[8-10], IL6[11], IL2R[11], galectin-1[12], and soluble CD30[13]. Importantly, these studies did not detail the particular cell subsets that are altered in cHL. Characterization of immune cell subsets in lymphoma is important, because peripheral perturbations can influence anti-tumor immunity, treatment responses, or immunity to other diseases. Furthermore, the development of prognostic- and treatment-related biomarkers will be most efficient when candidates are identified from peripheral blood, since this sample type is amenable to routine monitoring. A first step in biomarker development is to characterize the systemic immune cells present in cHL patients.

In recent years, an enormous variety of high parameter, single cell technologies have emerged to characterize immunophenotypes. High parameter flow cytometry is a

particularly well-established and robust platform for such work[14]. There are, however, major limitations in data analysis algorithms, which are often arcane, require a great deal of preliminary manual data analysis, or fail to comprehensively mine datasets[14]. Better data analysis tools will catalyze immuno-oncology and immunotherapy research, facilitating the identification of novel targets, and personalized medicine. In this paper, we introduce TerraFlow, a new data analysis platform that exhaustively identifies the cell populations whose frequency differs across the patient groups in a study.

Compared to classical supervised methods, which are limited to examining a few cell types (like naive, central memory, and effector T-cells), we demonstrate that TerraFlow can capture >150,000 cell phenotypes from a dataset. This is done using "combinatorics," a process by which cell populations are defined by the combinations of proteins they express[15]. TerraFlow improves this process through a number of novel advances. It streamlines the generation of cell phenotypes – producing combinations of six markers at a time – instead of the full set of 3^x phenotypes (where x = number of parameters measured). We developed this approach because the full set of combinatorial phenotypes includes many rare cell populations, defined by long and complex marker sets. This "phenotype inflation" is computationally intensive and creates severe statistical penalties when adjusting for false discovery rate (FDR). Since TerraFlow does not generate inordinately complex phenotypes, it is easier to identify the simple phenotypes that also drive cell subset/disease associations.

Additionally, TerraFlow offers an optional method that eliminates the need to create thresholds ("gates") that distinguish cells positive and negative for a marker, greatly simplifying flow cytometry data analysis; this method analyzes data on a sigmoidally-transformed continuous scale. TerraFlow also simplifies interpretation of results by trimming "redundant" markers, grouping cell phenotypes into families that are associated with outcome, and identifying the minimum set of markers that drives the difference between patient groups. Each of these features offers an advance over current methods for immune monitoring and biomarker discovery.

Using the power of high-parameter single-cell technology, and our new data analysis approach, we investigated whether newly-diagnosed cHL was associated with perturbations in functional features of the T-cell compartment, and whether these perturbations resolved after treatment. Specifically, we used high-parameter flow cytometry to perform direct *ex vivo* measurements of cellular proteins associated with activation, exhaustion, and suppression of T-cells in the peripheral blood. In addition, we studied cytokine expression (after *in vitro*, polyclonal re-stimulation) in the context of T-cell differentiation, exhaustion, and activation. Our results catalogue, with unprecedented depth, the nature of peripheral immunity in cHL patients, revealing previously unappreciated systemic immune abnormalities in cHL patients.

METHODS

Human Subjects. Informed consent was obtained from cHL patients (N=44) treated at the Perlmutter Cancer Center at NYU Langone Health and New York Presbyterian Weil Cornell (NYP) between 2011 and 2016. Patients with active viral infection or autoimmune disease were excluded. Blood samples were drawn pre-treatment and at sequential time points during and after therapy, and then processed for viable cryopreservation of peripheral blood mononuclear cells (PBMC); for this study, PBMC taken before treatment and at the three-month follow-up time point (after completion of therapy) were used. Age-matched, cryopreserved healthy donor PBMC (n=25) were obtained from STEMCELL Technologies (Cambridge, MA). Patient numbers for each analysis in the manuscript were as follows: Newly-diagnosed vs. Healthy: Checkpoint Panel - 30 cHL patients vs. 25 healthy donors, Cytokine Panel – 28 cHL patients vs. 25 healthy donors; Post-treatment vs. Healthy: Checkpoint Panel – 25 cHL patients vs. 25 healthy donors, Cytokine Panel – 21 cHL patients vs. 25 healthy donors. The number of patients studied was based on solely on sample availability; one cell vial was used for the Checkpoint Panel, a second (when available) was used for the Cytokine Panel. The median age for HL patients was 34.5, as is typical of this disease, but the age range was wide from 18-90 years (52% male, 48% female). Eighty percent of cHL patients had nodular sclerosing histology, 10% of patients had mixed-cellularity histology, and 3% had lymphocyte-rich histology, two patients had unspecified histology. Most patients (64%) had stage II disease, followed

by stage IV disease (21%), and stage III disease (14%; no patients had stage I disease). For post-treatment analyses, patients had received ABVD +/- consolidative radiation.

Cell Processing and High Parameter Flow Cytometry. Patient-derived whole blood was processed for isolation of PBMC using density-gradient centrifugation, and cells were re-suspended at a concentration of 10 million cells/mL for viable cryopreservation at -135C. For the experiments described here, samples were thawed, washed in RPMI media (Invitrogen, Carlsbad, CA) and split into two aliquots of equal cell numbers. The first aliquot was stained with a panel of antibodies described in previous work[15,16], while the second was stimulated for four hours with phorbol myristate acetate (PMA) and ionomycin (Sigma, St. Louis, MO), in the presence of Golgi Plug (containing Brefledin A; BD Biosciences, San Jose, CA). At the end of the stimulation period, the second aliquot of cells was stained with the flow cytometry antibodies described in **Supplemental Table 1**. Samples were analyzed immediately on a Symphony Flow Cytometer (BD Biosciences, San Jose, CA). Flow cytometry staining from a representative patient is shown in **Supplemental Figure 1A.**

Sample Analyses. Our analytical pipeline consists of two parts: feature engineering and feature selection. This first generates every possible phenotype consisting of 1 to N markers, using traditional Boolean gates or a novel non-gating approach. This combinatoric feature set will include many more phenotypes than needed to establish association with clinical outcome, including overlapping populations and extraneous markers. To resolve these redundancies, features are arranged into a network that captures the gradient changes that occur as markers are successively added, removed, and swapped between related phenotypes. The structure is used to derive a small subset of ”core nodes” that represent distinct trends in the data. These are further organized into families if they occupy overlapping regions in single-cell space.

Co-expression Weighting. In traditional flow cytometry analysis, populations are delineated by drawing bounding boxes over two-dimensional scatterplots. Cells inside boundaries are assigned a weight of 1 and those outside a weight of 0. The mean weight

equals some population frequency f. Co-expression weighting extends this process by replacing Boolean values with a continuous measure of co-expression. Cells are weighted by the cumulative sigmoidal product of each marker p and the average weight is returned as the overall amount of co-expression in n-dimensional space (for example, $\bar{p}$\{CD4+CD8+\}. The sigmoidal transformation ensures that double-positive cells have an exponentially larger weight than single- and double-negative cells, allowing rare populations to affect the overall mean. For negative gates, the fluorescence intensity distribution is simply reversed.

$f{+}{+} = \{CD4 \geq tCD4\} \cap \{CD8 \geq tCD8\}$ (Boolean)
$p{+}{+} = \sigma(CD4) \cdot \sigma(CD8)$ (Nongating)

Network Analysis. The combinatoric set contains many more features than are needed to establish association with clinical outcome. To resolve redundancy between similar phenotypes, features are arranged into a network that captures global trends in the data. The process begins by calculating the correlation of each phenotype with clinical outcome. We used the coefficient of determination but other zero-centered measures such as log-ratio are appropriate as well. Next, features are arranged into a network by adding edges where phenotypes differ by the addition or removal of one marker. For example, one path through the network might pass through nodes {CD95+} → {CD4+CD95+} → {CD4+CD8-CD95+}. Finally, each node is queried to determine if its correlation is stronger than any node adjacent to its position. These local maxima or ”core nodes” are selected to represent their region of the network. If necessary, a smoothing effect can be applied by averaging the effect size of each node with the average effect size of its neighbors.

Selected phenotypes are then queried again, this time to determine if any markers can be removed without significantly compromising correlation to clinical outcome. If a simpler version of the phenotype has an effect size within tolerance of the optimal phenotype (we used 95%), the extra marker is removed. For instance, {CD4+CD8−} has a correlation of $r = 0.9$ but {CD4+} has a correlation of $r = 0.89$. In this case, the extra CD8− marker would

be dropped in favor of a simpler representation. The process is repeated until all of the markers in the final feature set are deemed nonredundant.

Cluster Analysis. Despite being selected against related phenotypes, the final feature set may still delineate overlapping cell populations. To resolve these redundancies, cell weights are calculated for each combinatoric in the final feature set and compared using Jaccard or Spearman distances. Phenotypes that frequently appear on the same cells are grouped using hierarchical clustering.

Recursive Feature Elimination. Whereas network analysis enumerates all differentially expressed phenotypes, recursive feature elimination (RFE) determines how few phenotypes and markers provide the same level of accuracy as models trained on the entire dataset. Markers are sequentially removed from the data set and logistic regression models are fit to the remaining predictors. Classification accuracy is estimated using leave-one-out cross-validation and the marker with the smallest effect on performance is dropped. The process repeats until there is only one marker remaining.

Weighted Lasso. Each regularized regression algorithm has a different way of dealing with multicollinearity. In Elastic Net regression, collinear predictors are retained as a set. This may be useful for exploratory purposes but will end up selecting more predictors than necessary to achieve the same level of accuracy. In Lasso regression, one predictor is selected from each collinear set. The choice of predictor, however, is difficult to control. All things being equal, a 17-marker phenotype is just as likely to be selected as a 2-marker phenotype because the algorithm doesn't incorporate metadata such as phenotypic complexity. We used Lasso regression with an additional penalty term that increases with number of markers per phenotype. The penalty acts as a tiebreaker and forces the algorithm to choose the simplest phenotype from each collinear set. Combinations of three and more markers can be included in the model, but only if they provide a larger boost in performance than their lower-order counterparts. This approach was used to identify a minimal set of low-order phenotypes whose linear combination yields accurate predictions of clinical outcome.

Data Interpretation. A new platform developed by Epistemic AI (New York, NY) for mining scientific literature was used to aid in the understanding of cell phenotypes identified.

RESULTS

TerraFlow: A New High Parameter Data Analysis Pipeline

Figure 1A illustrates TerraFlow features (colored boxes), alongside existing approaches (white boxes). **Figure 1B** shows how Boolean combinations of markers are constructed into cell phenotypes, using flow cytometric gating. This approach assigns a value of zero to any marker/cell event that is below the defined threshold, and a value of one to cell events above the threshold. Non-gating combinatorics extends this cell weighting process by replacing Boolean values with a continuous scale. Cells are weighted by the sigmoidal product of each marker (**Figure 1C**) and the average weight is returned as the overall amount of expression for that sample. The sigmoidal transformation ensures that double-positive cells have an exponentially larger weight than single- and double-negative cells, allowing even rare events to affect the overall mean.

To define the performance of the non-gating approach, we compared weighting in our non-gating method to the traditional Boolean gates (**Supplemental Figure 1B**). Cells expressing a particular phenotype, as defined by Boolean gating, had larger cell weights than cells that were not of that phenotype, demonstrating that our approach is relatively consistent with classical methodologies. Notably, it appears that our non-gating method captures more variation within a gate (**Supplemental Figure 1B)**, as cell weights vary more in positive than negative gates. **Figure 1D** further demonstrates concordance between the approaches at the population level, illustrating the strong correlation between population frequencies defined by non-gating and gating methods. Correlation, however, is reduced in an exceedingly infrequent cell population, defined by four cell surface markers (**Figure 1D**). This phenomenon generalizes to all rare cell populations in our study, as the correlation between phenotypes defined by each approach is higher when the number of markers involved is low (**Supplemental Figure 1C**). This is an expected

feature of non-gating combinatorics, since higher-order combinatorial phenotypes are more likely to describe less frequent populations. The phenomenon weeds out very rare cell populations from analysis reports, so that investigators can focus on those that are mostly easily measured. (For users interested in capturing very rare cell types, TerraFlow still offers an option to use the traditional combinatorics approach, which is not biased against very rare cell types.)

Additional features, benchmarking, and results of TerraFlow are described below, in the context of the clinical research questions we studied. First, we asked whether T-cell function (as defined cell activation, exhaustion, and cytokine production) are impaired in newly-diagnosed cHL patients, compared to healthy controls. Next, we asked whether any functional differences emerged or persisted after treatment.

Mapping the topology of T-cell function in newly-diagnosed cHL patients

Dimension reduction provides only limited information. Data from two high-parameter flow cytometry panels was analyzed using common dimension reduction and cell population identification algorithms (UMAP and DBSCAN). The first panel (Checkpoint) measured markers of cell differentiation, activation, trafficking, and exhaustion, while the second panel (Cytokine) measured expression of a number of T-cell associated cytokines, in the context of cell differentiation, activation, and exhaustion. For each panel, UMAP/DBSCAN identified a single cell type that significantly differed between newly-diagnosed and healthy patients for each panel (Checkpoint: Cluster 3, CD8+ CD127+ CCR7+ CD45RO- CD95- TIGIT- HLADR- PD1- CD278- CD103- CD152- GITR- CD366- CD69- CD57-, $p<0.05$ (**Figure 2A**); Cytokine: Cluster 3, CD8+ CD278+ CCR7+ CD95- CD45RO- CD57- TNF- IFNG- PD1- IL17- CD152- CD134- IL4- CD366- IL13-, $p<0.05$ (**Supplemental Figure 2A**)). Each cell type was diminished in newly-diagnosed patients compared to healthy donors.

We further described these cells using a new platform for cell subset interpretation, built by Epistemic AI, which mines scientific literature and databases for molecule, cell type, gene network, pathway, process, and drug information. For simplicity, only non-lineage

markers that were expressed by these two cell populations were queried. CD127+ CCR7+ and CD278+ CCR7+ cell populations have not been previously reported for their association with newly-diagnosed cHL; however, CD127+ CCR7+ are commonly recognized in literature describing viral-specific CD8+ central memory T-cells. The Epistemic AI platform reports that these cells may engage IL7-receptor and G-alpha signaling pathways, through expression of CD127 and CCR7 respectively, leading to positive regulation of STAT, T-cell receptor, CCL21, and CCL19 signaling processes and IL12 and IFNg production. CD278 and CCR7 have been studied extensively in the context of follicular helper (CD4+) T-cells, but not in the context of peripheral blood expression amongst CD8+ T-cells. Epistemic AI also reports that these molecules are known to enhance T-cell stimulation (both CD278 and CCR7) and direct T-cell trafficking into the lymph node (CCR7). In sum, these findings suggest that newly-diagnosed cHL patients have reduced levels of early memory T-cells, and may have fewer cells that express proteins involved in amplification and maintenance of immune responses.

Notably, the small number of phenotypes identified by dimension reduction are defined by many markers, and dimension reduction cannot further indicate whether 1) there are other phenotypes that behave similarly (both related or unrelated to those reported), or 2) simpler phenotypes that drive the association. With this information, a fuller picture of the difference between newly diagnosed and healthy individuals would be available.

Gating-based combinatorics. To more comprehensively mine the data, we applied our new streamlined combinatorics approach, which enumerates cells expressing every possible combination of six markers at a time (Checkpoint: 109,916 cell types; Cytokine: 79,625 cell types). We then developed a simple scoring system that ranks the cell types by FDR-adjusted p-value and fold difference between newly-diagnosed patients and healthy donors, as well as median frequency, so that we could identify the most dramatic differences from amongst the many phenotypes studied.

For the Checkpoint panel, 586 cell populations were differentially observed between newly-diagnosed patients and healthy donors at an FDR-adjusted p-value<0.05.

Supplemental Data File 1 lists the phenotype of these cell populations, and various other metrics, including those used to score and rank the populations. **Figure 2B** shows the phenotype of the top 25 highest scoring phenotypes that are enriched in newly-diagnosed patients. Expression of CD45RO, CD272, and CD95 generally characterizes these populations, all of which lack expression of CD127. In healthy individuals, cells expressing CD127 are enriched (**Supplemental Data File 1**).

For the Cytokine panel, no cell populations were different between newly-diagnosed and healthy donors at FDR-adjusted p-value<0.05. However, 2,009 cell populations were differentially observed at FDR-adjusted p-values that trended toward significance (0.073 - 0.075). These are listed in **Supplemental Data File 2**. The highest scoring amongst these included cells expressing CD278, CD95, and PD1. These findings, as reported by the Epistemic AI platform, suggest that newly-diagnosed patients have elevated levels of memory (CD45RO+) and activated (CD278+) cells, primed for apoptosis (CD95+), and expressing the inhibitory/exhaustion receptors CD272 (BTLA) and PD1.

Non-gating combinatorics pipeline. We implemented our novel non-gating combinatorics as a pipeline that included network analysis (to identify families of related phenotypes). For the Checkpoint dataset, we randomly selected 44 training samples, and used non-gating combinatorics to generate 175,000 cell phenotypes, whose frequency was determined for each patient. We filtered these cell phenotypes to exclude those with fewer than 1000 events across all samples (i.e., ~20 cells/sample). This left 110,000 phenotypes, whose frequency was compared between newly diagnosed patients and healthy donors. There were 2789 cell populations that were significantly different between study groups (with an FDR-adjusted p-value$\leq$0.05, 1297 were elevated in newly-diagnosed patients, while 712 were elevated in healthy donors; **Figure 2C**).

Because the set of combinatorial phenotypes contained many overlapping and closely related cell subsets, we developed a “local maxima” methodology to identify core phenotypes. Features were arranged into a network that captured smooth gradients in effect size occuring as markers are swapped in and out of combinations (see *Network*

Analysis in **Methods**). By arranging phenotypes into a network and querying them against their nearest neighbors, we identified a small number of core nodes that captured distinct trends in the data. This process reduced the 2789 cell populations that varied across study groups to 213 representative, core phenotypes.

Selected phenotypes were then queried again, this time to determine if any markers could be removed without significantly compromising correlation to clinical outcome. If a simpler version of the phenotype had an effect size within tolerance of the optimal phenotype (we used 95%), the extra ("redundant") marker was trimmed away. The process (**Figure 2C**) was repeated until all of the markers in the final feature set were deemed non-redundant; for the Checkpoint dataset, this process yielded 140 cell phenotypes. For the Cytokine dataset, it yielded 194 cell phenotypes (**Supplemental Figure 2B**).

We returned to the single cell data and mapped each phenotype to a cell that expressed that particular combination of markers. Cells that consistently expressed the same phenotypes were identified and those phenotypes were grouped by hierarchical clustering, which defined families of phenotypes. For the checkpoint dataset, 77 families were identified; the mapping of phenotypes to families is reported in **Supplemental Data File 3,** and illustrated in **Figure 2D**. For the Cytokine dataset, 84 families were found (**Supplemental Data File 4**; heat map in **Supplemental Figure 2C**).

We next returned to the phenotypes resulting from the trimming step, and fit a simple linear regression model, which demonstrated a strong difference between newly-diagnosed and healthy individuals in the Checkpoint data training set ($p<2E-16$, **Figure 2E**). For a set of 11 samples held out for validation, this model had an area-under-the-curve (AUC) of 0.933, with a p-value of 0.024 describing the difference between newly-diagnosed and healthy study participants (**Figure 2E**). Similar performance was observed with the Cytokine data set, where the model describing the difference between the two patient groups had a p-value=7.2E-9 in the training set. The validation set reported an AUC of 0.933, with a p-value=0.018 (**Supplemental Figure 2D**).

The top 30 core phenotypes (as determined by p-value) were visualized using a force directed graph. Edges were weighted by the Spearman correlation coefficients among single-cell weights. For the Checkpoint dataset, newly-diagnosed patients (**Supplemental Figure 3A-B**) had three networks of related phenotypes, across eight family clusters, while healthy donors had three networks across five family clusters (**Supplemental Figures 3C-D**). For the Cytokine dataset, newly diagnosed patients (**Supplemental Figure 4A-B**) exhibited a single network that included eight family clusters, while healthy individuals exhibited four networks that contained six families of cell types.

For the most statistically significant core nodes identified in the Checkpoint dataset ($p<0.0001$), **Figure 3A** shows the differences between newly-diagnosed and healthy individuals. Our approach revealed that specific populations of cells expressing various combinations of immune checkpoints and regulatory markers, like CD95, CD152, CD272, CD278, and GITR, are elevated in newly-diagnosed individuals. This data reveals global systemic T-cell exhaustion and dysregulation in newly diagnosed cHL patients, along with the specific cell phenotypes involved, and may suggest novel combinations of immunotherapy agents that could target cell types elevated in cHL. In healthy individuals, specific cell phenotypes expressing CD127 are enriched, suggesting that cHL is also associated with a loss of CD127-expressing naive and early memory cells, and that premature immune ageing and senescence is observed in the T-cell compartment of cHL patients.

For the most statistically significant core nodes in the Cytokine dataset ($p<=0.001$), **Supplemental Figure 5A** shows that differences between newly-diagnosed and healthy individuals. Cell populations expressing combinations of CD57, CD95, and CD278 are elevated in cHL patients, as well as cells expressing CD366 and cells that make IL17. In healthy individuals, cells expressing IL4 or IFNg are more prevalent. These data provide information complementary to the results of the Checkpoint analysis; they highlight immune exhaustion and a somewhat diminished cytokine effector function in newly-diagnosed cHL patients.

Recursive feature elimination (RFE). While network analysis aims to summarize *all* differentially expressed phenotypes, RFE seeks to determine the smallest number of markers and phenotypes that can achieve the same accuracy as models trained on the entire dataset. Thus, our RFE algorithm can build the simplest models to describe the differences between patient groups. RFE builds models using data from the gating-based and non-gating combinatorics, tests both elastic net and weighted lasso machine learning algorithms, and examines performance of models that use progressively fewer markers. This screening step is done on a training set of data, and then the best model is re-trained on the entire training set and performance is validated on a holdout dataset. Although all pairs are tested and reported by TerraFlow, the combination of non-gating combinatorics and weighted lasso machine learning is preferred when performance between the approaches is similar, because of the non-gating approach's simplicity and weighted lasso's use of less complex phenotypes.

For the checkpoint dataset, all models screened (pairings of the two combinatorics methodologies with the two machine learning algorithms) had similar performance (data not shown). **Figure 3B** shows the elimination path for the non-gating/weighted lasso model, which captures the relative importance of CD366, CD278, CD152, CD69, TIGIT, and CD8 in defining the differences between newly-diagnosed patients and healthy donors. Indeed, the machine learning model that incorporated these markers exhibited a remarkable AUC of 0.967, with strong significance (training set: $p<2E-16$; test set: $p=0.0038$). Analysis of feature importance (**Figure 3C**) shows that enumeration of CD152+ CD278+ is most important in finding this difference, and **Figure 3D** reports the difference of each feature alone (without measurement of other cell types) between patient groups. In sum, this process identifies a minimal set of markers that can be measured instead of the full Checkpoint panel, builds a machine learning model that best identifies group differences using those core phenotypes, and then reports that CD152+ CD278+ cells are a major, but not sole driver, of the difference between study groups.

For the Cytokine dataset, models that used gating-based combinatorics outperformed those that used the non-gating approach, in terms of machine learning results (data not shown). **Supplemental Figure 5B** shows the elimination path for the weighted lasso machine learning algorithm, which highlights the relative importance of measurements of TNF, IFNg, CD366, CD152, and PD1 in the Cytokine antibody panel. The model built validated well, with an AUC of 0.9, and statistically-significant differences between newly-diagnosed and healthy individuals (training set: p=2.3E-7; test set: p=0.024; **Supplemental Figure 5C**). Analysis of feature importance revealed CD366+ and CD366+ TNF+ cell populations as major drivers of the difference (**Supplemental Figure 5D**); indeed, most newly-diagnosed individuals had detectable levels of CD366+ TNF+ cells, while healthy individuals did not (**Supplemental Figure 5E**; p<0.001).

Post-treatment differences with healthy individuals

We also used TerraFlow to characterize differences in checkpoint molecule and cytokine expression between patients after treatment and healthy donors. Dimension reduction, using UMAP and DBSCAN, revealed a few cell types elevated in healthy individuals (**Supplemental Figure 6A-B**). However, our gating-based combinatorics approach identified 1,243 and 9,909 cell phenotypes that differed significantly in the Checkpoint and Cytokine datasets, respectively (FDR-adjusted p-value<0.05). Amongst the 1,243 populations reported by analysis of the Checkpoint dataset, 1,189 were diminished and 54 were elevated in patients post-treatment (compared to age-matched healthy donors). Amongst the 9,909 cell phenotypes reported from the Cytokine dataset, 2,526 were reduced and 7,383 were elevated in patients post-treatment. These populations, and the p-values of the associated comparisons between post-treatment patients and healthy donors, are reported in **Supplemental Data File 5** (Checkpoint panel) **and Supplemental Data File 6** (Cytokine panel). For the Checkpoint dataset, the phenotypes with the highest score (as defined by the frequency, p-value, and fold-change rankings) generally expressed combinations of CD272, CD45RO, and CD95; these were elevated in patients post-treatment, compared to healthy controls. For the Cytokine dataset, the highest scoring phenotypes CD278 and CD95; these were enriched in patients post-treatment. Taken together, these phenotypes suggest that specific populations of

exhausted, activated cells that are primed for apoptosis are observed in patients post-treatment.

Our non-gating combinatorics pipeline created 175,000 and 165,000 phenotypes for the Checkpoint and Cytokine datasets, respectively (**Figure 4A, Figure 5A**). Of these phenotypes, 1792 and 6792 were present at different frequencies in post-treatment and healthy individuals, respectively (FDR-adjusted p-value<0.01, Checkpoint; p-value<0.005, Cytokine). Subsequent local maxima analysis and trimming for redundant markers yielded 181 and 91 core phenotypes for the Checkpoint and Cytokine datasets, respectively.

The logistic regression models that best described post-treatment/healthy donor differences in the Checkpoint dataset yielded excellent performance (**Figure 4A;** AUC=1; p<2E-16 (training set); p=9.5E-5 (test set)). Three networks were observed that described cells enriched in healthy donors, which included eight cell population clusters of 103 cell families (**Figure 4B**). For cells enriched in post-treatment subjects, three networks were also observed, incorporating six cell population clusters. Most of the core phenotypes that were reduced in patients post-treatment (compared to healthy controls) were CD127+ cells, occasionally also expressing PD1 (p<0.0001, **Figure 4C**). This suggests that the loss of naive and early memory cells in newly-diagnosed patients persists after treatment. Patients also exhibit higher levels of CD272+, GITR+, CD366+, TIGIT+, and CD278+ cells (p<0.0001), suggesting that immune exhaustion persists after treatment.

The logistic regression models for the Cytokine dataset also perform well (**Figure 5A**; AUC=1; p=1.1e-12 (training set) and p=0.008 (test set)). We observed three networks of cells enriched in healthy donors (**Figure 5B**), but the cell types enriched in post-treatment patients appear to be contained in the same network. The vast majority of core phenotypes reduced in cHL patients post-treatment are cell types that produce IFNg (p<0.0001; **Figure 5C**), suggesting a deficit in production of this important T-cell effector cytokine. Core phenotypes that are elevated in post-treatment patients express combinations of CD278, CD57, CD95, CD366, CD152, and IL17 (**Figure 5C**). Again,

these results suggest that the immunologic abnormalities in newly-diagnosed patients are not normalized after treatment.

Finally, our RFE algorithm reported a simple set of phenotypes that represents post-treatment abnormalities. For the Checkpoint dataset, non-gating combinatorics outperformed Boolean combinatorics, with both weighted lasso and elastic net algorithms performing well. The elimination path (**Figure 6A**) reveals the importance of CD278, CD366, CD272, and CD127 (**Figure 6B**) in constructing a machine learning model for post-treatment differences (model AUC=1; p<5.9e-12 (training set); p=0.0066 (test set); **Figure 6C-D**). Amongst the cell types expressing these markers, CD272+ CD366+ cells are most important; these cells are elevated in post-treatment donors (p=1.2E-6; **Figure 6E**). For the Cytokine dataset, the elimination path included CD95 and CD366 (**Figure 6F**), and - as observed with the Checkpoint dataset - cells expressing CD366 represented the most important feature (**Figure 6G**). For machine learning identification of post-treatment immune abnormalities, the non-gating/weighted lasso pair performed best (AUC=1, **Figure 6H**), with clear distinction between post-treatment patients and healthy donors (p=2.6E-8 (training set) and p=0.054 (test set), **Figure 6I**). CD366+ cells were elevated in post-treatment patients compared to healthy donors (p=3.4E-7, **Figure 6J**).

DISCUSSION

Our study reveals a number of perturbations to the T-cell compartment in cHL patients. These include higher levels of activated (CD278+), exhausted (CD366+, PD1+, CD152+), and suppressive (GITR+) T-cells compared to healthy donors, and diminished levels of T-cells producing effector cytokines (like IFNg and IL4). Subsets of cytokine-producing cells that co-express markers of exhaustion (i.e., TNF+ CD366+ cells) are also elevated in cHL patients. Finally, T-cells expressing CD127 are reduced in cHL patients; this molecule is a receptor for IL7, which is involved in homeostatic renewal of cells and is typically observed on naive and central memory T-cells. Taken together, these findings suggest that the systemic T-cell compartment in cHL patients is shifted away from less differentiated cells, with the potential for self-renewal, toward a more exhausted profile.

T-cell effector function also appears reduced in cHL patients, and our data suggest a shift away from T-helper 1 and T-helper 2 type responses toward T-helper 17 cells. Many of these immune perturbations remain even after treatment.

Our findings provide mechanistic explanations for historical reports of immune deficiency in cHL patients. In the 1960s and 1970s, a series of studies demonstrated impaired cell-mediated immunity in cHL, including anergy to recall skin tests[17-19], failure of delayed-type hypersensitivity (DTH) reactions[20], prolonged survival of skin allografts[19,21], and impaired response to BCG vaccination[22]. Potential mechanisms for these defects include the immune exhaustion and shift away from Th1 cytokine responses that we observed.

More importantly, however, our detailed inventory of the cells elevated in cHL patients provides evidence to support the use of new immunotherapy agents directed against CD366[23] or GITR[24], for example, and may inform design of combination immunotherapy strategies that target multiple markers elevated in cHL patients. We are currently using TerraFlow to determine which specific cell phenotypes are associated with therapeutic failure, with the aim of contributing to biomarker development and individualized (precision oncology) treatment strategies. We demonstrate here that our comprehensive approach can discover many disease-relevant cell types in the peripheral blood, and distill that enormous list of phenotypes into populations defined by simple marker sets. Cells defined by simple marker sets are more suitable for validation/biomarker development, can be measured with less complex and more inexpensive technology, and have the potential for widespread application. Currently, candidate biomarkers for patient risk or clinical outcome in cHL require measurements of rare HRS cells or invasive sampling of the tumor microenvironment[25]; however, with our introduction of a comprehensive method to analyze data from peripheral blood, candidate biomarkers from this more accessible sample type may soon emerge.

TerraFlow is unique for the depth to which it mines high-parameter, single cell datasets. Our method represents a step forward from the common supervised analysis approach, which tracks only a limited number of canonical cell subsets defined by only a few

markers. We demonstrate a dramatic advantage over the most common unsupervised data analysis approach, dimension reduction coupled with clustering, which often results in ambiguous clusters and only identifies a few (if any) disease-related cell types. TerraFlow, in contrast, allows examination of 150,000 to 200,000 cell phenotypes for high parameter cytometry panels (<30 markers).

Our system is designed to avoid potential problems and enhance usability. The first major bottleneck in combinatorics is the vast amount of redundancy in feature seta. Every true effect in the data may be reflected in hundreds of phenotypes that differ by one or two markers, complicating efforts to select features using traditional methods that assume independence among predictor variables. Here, we developed a more rational feature selection method that incorporates metadata such as number of markers and amount of overlap between related phenotypes. This network approach allows us to identify families of phenotypes that represent distinct trends in the data.

Whereas the network approach seeks to identify *all* differentially expressed phenotypes, a separate segment of the pipeline selects a *minimal* set of markers and phenotypes whose linear combination can be used to achieve accurate classification of clinical groups. This may be useful for early exploratory analysis and for the design of simpler follow-up experiments.

The second major bottleneck is the need to manually threshold positive and negative events, a time-consuming process that may obscure continuous gradients in expression that have biological meaning. We introduce a simple mathematical model that allows us to measure coexpression on a continuous scale, obviating the need for manual gating and allowing us to detect finer nuances in differential expression. As the number of markers that can be studied in single-cell, high parameter experiments increases this time-saving step will become increasingly valuable. In conclusion, TerraFlow catalogued systemic T-cell exhaustion and dysregulation of cytokine production in cHL. The features of TerraFlow are well-suited to analyze the modern generation of complex, high-parameter datasets.

ACKNOWLEDGEMENTS

The authors would like to thank Suraj Saksena for coordinating the BD Bioscieces/Precision Immunology Laboratory visiting scientist program under which TL began these experiments. We would also like to thank additional investigators who contributed patients to this study: Tibor Moscovits and Kenneth Hymes (NYU Langone Health), as well as Peter Martin and John Leonard (Weill Cornell). We also acknowledge funding support from the Feinberg and Greenberg families. CD was supported by American Cancer Society grant: MRSG-14-052-01-LIB.

AUTHOR CONTRIBUTIONS

CD conceived of and designed studies, led the clinical study and recruited patients, analyzed and interpreted data, and wrote manuscript. DF performed experiments, developed TerraFlow, analyzed data, and wrote manuscript. TL performed experiments. LL organized study and performed experiments. JA organized study and performed experiments. BR, MG, DK, and JR conducted the clinical study and recruited patients. PKC conceived of and designed studies, developed and benchmarked TerraFlow, analyzed and interpreted data, and wrote manuscript.

FIGURE LEGENDS

Figure 1. *TerraFlow Background and Performance.* A) Approaches for flow cytometry data analysis, pictured within the context of TerraFlow. Older methodologies are presented in white boxes, colored boxes indicate elements and features of TerraFlow. The value proposition underlying each method is indicated. B) Construction of phenotypes from the gates for each marker. Marker expression for each patient is depicted with gray lines; average expression, from which gates are derived, is shown in blue. Examples of combinatorial phenotypes are shown. C) Bivariate flow cytometry plots with yellow events depicting cells within a standard flow cytometry gate (left). In contrast,

the sigmoidal transformation used in our non-gating routine captures a continuous scale of expression (from red to orange to yellow events). D) Correlation between the frequency and weight of cells defined by Boolean gating and non-gating combinatorics, respectively.

Figure 2. *TerraFlow Comparison of Newly-Diagnosed cHL Patients and Healthy Donors (Checkpoint Dataset).* A) UMAP projections of data from newly-diagnosed cHL patients (left), healthy donors (middle), and an overlay of clusters identtified by DBSCAN (right). Heat map shows expression level of each marker (columns) in each cluster (rows) for newly-diagnosed patients. Line and circle plot depicts difference between study groups as effect size; colored circle identifies cluster with statistically significant difference. B) Output from gating-based, streamlined combinatorics approach. Phenotyping heat map of top 25 highest scoring populations that differed between newly-diagnosed cHL patients and healthy donors; scoring based on rankings of FDR-adjusted p-value, median frequency, and fold difference. Red boxes indicate positive expression for a marker, black boxes indicate lack of expression, clear boxes indicate markers that are not considered in phenotype. C) Summary of screening steps performed by TerraFlow using non-gating approach; numbers indicate the number of phenotypes remaining after each step. D) Heat map showing 77 clusters of phenotypes ("families" from panel C) that are significantly different between newly-diagnosed and healthy donors (coloring indicates effect size of phenotype). E) Performance of a linear regression model that incorporates information from the 140 trimmed phenotypes (panel C).

Figure 3. *Specific Phenotypes Identified by TerraFlow and Their Reduction to Simplest Components Using Recursive Feature Elimination.* A) Core nodes (i.e., phenotypes), as defined by network analysis, are compared between newly-diagnosed and healthy donors. B) Recursive feature elimination shows the effect of removal of each marker (x axis) on the performance of the model (y-axis). C) Feature importance results from a weighted lasso machine learning model that incorporates phenotypes with the reduced marker set. The level of CD152+ CD278+ cells appears to be a major driver of the difference between newly-diagnosed and healthy donors. D) Significant phenotypes from the reduced model.

Figure 4. *TerraFlow Comparison of cHL Patients Post-Treatment and Healthy Donors (Checkpoint Dataset).* A) Summary of screening steps performed by TerraFlow using non-gating approach; numbers indicate the number of phenotypes remaining after each step (left panel). Performance of a linear regression model that incorporates information from the 181 trimmed phenotypes (center and right panels). B) Graphic showing networks of phenotypes associated with healthy donors (blue) and post-treatment cHL patients (orange and red). C) Core nodes (i.e., phenotypes), as defined by network analysis, are compared between post-treatment patients and healthy donors.

Figure 5. *TerraFlow Comparison of cHL Patients Post-Treatment and Healthy Donors (Cytokine Dataset).* A) Summary of screening steps performed by TerraFlow using non-gating approach; numbers indicate the number of phenotypes remaining after each step (left panel). Performance of a linear regression model that incorporates information from the 181 trimmed phenotypes (center and right panels). B) Graphic showing networks of phenotypes associated with healthy donors (blue) and post-treatment cHL patients (orange and red). C) Core nodes (i.e., phenotypes), as defined by network analysis, are compared between post-treatment patients and healthy donors.

Figure 6. *Recursive Feature Elimination on Phenotypes Differing Between Post-Treatment Patients and Healthy Donors.* A) The effect of removal of each marker (x axis) on the performance of the model (y-axis) for the Checkpoint dataset. B) Performance of the weighted lasso machine learning model that incorporates phenotypes from the reduced marker set. C) Feature importance results show that the level of CD366+ CD69- cells is a major driver of the difference between post-treatment patients and healthy donors. D) For the Cytokine dataset, the combination of the non-gating and weighted lasso machine learning approaches best captured the difference between patient groups; AUC analysis and training/test set results are shown for that pair of analysis approaches. E) The effect of removal of each marker (x axis) on the performance of the model (y-axis) for the Cytokine dataset. F) Feature importance shows that CD366+ cells are a major

driver of the study group differences. G) Level of CD366+ cells is shown for each patient group.

Supplemental Figure 1. A) Representative flow cytometry staining for Checkpoint (top) and Cytokine (bottom) panels. B) Comparison between non-gating weight and Boolean (positive vs. negative) categorization of marker expression level. C) Correlation between non-gating and Boolean approaches is associated with the number of markers used to define the phenotype, showing that non-gating approach has a useful bias against complicated phenotypes.

Supplemental Figure 2. A) Results from comparison of newly-diagnosed to healthy donors using dimension reduction, as perforrmed on the Cytokine dataset. B) Summary of number of phenotypes left after each step in the non-gating TerraFlow pipeline for Cytokine dataset. C) Heatmap depicting clustering of phenotype families and the effect size describing the difference between newly-diagnosed and healthy donors for each phenotype in the Cytokine dataset. D) Performance of a linear regression model that incorporates information from the 194 trimmed phenotypes in the Cytokine dataset.

Supplemental Figure 3. A) Three networks of related cell phenotypes are enriched in newly-diagnosed patients (compared to healthy donors) in the Checkpoint dataset. B) The three networks from panel A include eight families of phenotypes. C) Three networks of related cell phenotypes that are enriched in healthy donors (compared to newly-diagnosed patients) in the Checkpoint dataset. D) These three networks incorporate five families of phenotypes.

Supplemental Figure 4. A) One network of related cell phenotypes is enriched in newly-diagnosed patients (compared to healthy donors) in the Cytokine dataset. B) This network includes eight families of phenotypes. C) Four networks of related cell phenotypes that are enriched in healthy donors (compared to newly-diagnosed patients) in the Cytokine dataset. D) These four networks incorporate six families of phenotypes.

Supplemental Figure 5. A) Core nodes (i.e., phenotypes), as defined by network analysis, are compared between newly-diagnosed and healthy donors for the Cytokine dataset. B) Recursive feature elimination shows the effect of removal of each marker (x axis) on the performance of the model (y-axis). C) Feature importance results from a weighted lasso machine learning model that incorporates phenotypes with the reduced marker set. The level of CD366+ TNF+ cells appears to be a major driver of the difference between newly-diagnosed and healthy donors. D) Significant phenotypes from the reduced model.

Supplemental Figure 6. A) Dimension reduction is used to compare post-treatment patients and healthy donors for the Checkpoint dataset. B) Dimension reduction is used to compare post-treatment patients and healthy donors for the Cytokine dataset.

REFERENCES

1 Steidl, C., Connors, J. M. & Gascoyne, R. D. Molecular pathogenesis of Hodgkin's lymphoma: increasing evidence of the importance of the microenvironment. *J Clin Oncol* **29**, 1812-1826, doi:10.1200/JCO.2010.32.8401 (2011).

2 Aldinucci, D., Gloghini, A., Pinto, A., De Filippi, R. & Carbone, A. The classical Hodgkin's lymphoma microenvironment and its role in promoting tumour growth and immune escape. *J Pathol* **221**, 248-263, doi:10.1002/path.2711 (2010).

3 Kuppers, R. The biology of Hodgkin's lymphoma. *Nat Rev Cancer* **9**, 15-27, doi:10.1038/nrc2542 (2009).

4 Hsi, E. D. Biologic features of Hodgkin lymphoma and the development of biologic prognostic factors in Hodgkin lymphoma: tumor and microenvironment. *Leuk Lymphoma* **49**, 1668-1680, doi:10.1080/10428190802163339 (2008).

5 Khan, G. Epstein-Barr virus, cytokines, and inflammation: a cocktail for the pathogenesis of Hodgkin's lymphoma? *Exp Hematol* **34**, 399-406, doi:10.1016/j.exphem.2005.11.008 (2006).

6 Aldinucci, D. *et al.* Expression of CCR5 receptors on Reed-Sternberg cells and Hodgkin lymphoma cell lines: involvement of CCL5/Rantes in tumor cell growth and microenvironmental interactions. *Int J Cancer* **122**, 769-776, doi:10.1002/ijc.23119 (2008).

7 Tanijiri, T. *et al.* Hodgkin's reed-sternberg cell line (KM-H2) promotes a bidirectional differentiation of CD4+CD25+Foxp3+ T cells and CD4+ cytotoxic T lymphocytes from CD4+ naive T cells. *J Leukoc Biol* **82**, 576-584, doi:10.1189/jlb.0906565 (2007).

8 van den Berg, A., Visser, L. & Poppema, S. High expression of the CC chemokine TARC in Reed-Sternberg cells. A possible explanation for the characteristic T-cell infiltratein Hodgkin's lymphoma. *Am J Pathol* **154**, 1685-1691, doi:10.1016/S0002-9440(10)65424-7 (1999).

9 van Rijswijk, R. E., de Meijer, A., Sybesma, J. P. & Kater, L. Five-year survival in Hodgkin's disease. The prospective value of immune status at diagnosis. *Cancer* **57**, 1489-1496, doi:10.1002/1097-0142(19860415)57:8<1489::aid-cncr2820570809>3.0.co;2-n (1986).

10 Weihrauch, M. R. *et al.* Elevated serum levels of CC thymus and activation-related chemokine (TARC) in primary Hodgkin's disease: potential for a prognostic factor. *Cancer Res* **65**, 5516-5519, doi:10.1158/0008-5472.CAN-05-0100 (2005).

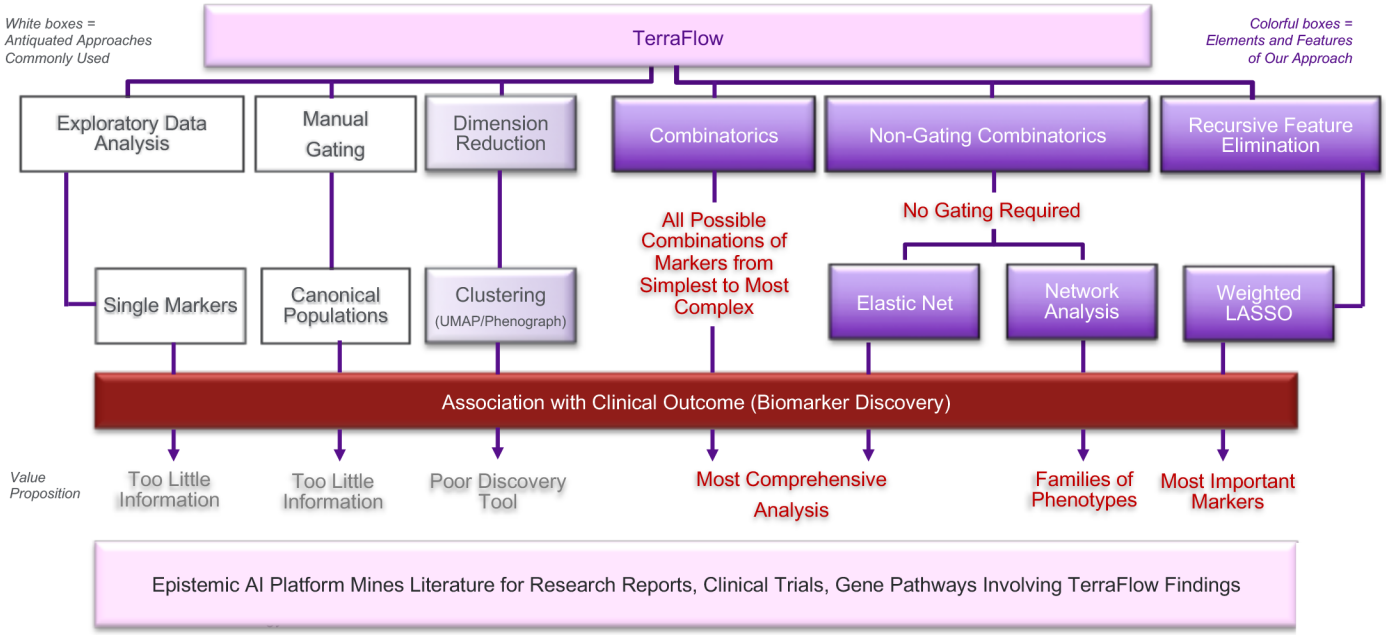
11 Ansell, S. M. Annual clinical updates in hematological malignancies: a continuing medical education series. Hodgkin lymphoma: 2011 update on diagnosis, risk-stratification, and management. *Am J Hematol* **86**, 851-858, doi:10.1002/ajh.22105 (2011).

12 Ouyang J, P. A., Pogge E, Ponader S, Rabinovich G, Neuberg DS, et al. Galectin-1 Serum Levels Reflect Tumor Burden and Adverse Clinical Features in Hodgkin Lymphoma. *ASH Annual Meeting Abstracts* **120**, 1 (2012).

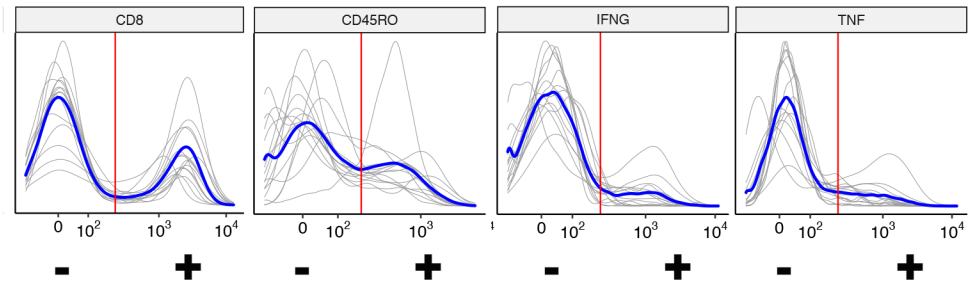
13 Levin, L. I. *et al.* Elevated Serum Levels of sCD30 and IL6 and Detectable IL10 Precede Classical Hodgkin Lymphoma Diagnosis. *Cancer Epidemiol Biomarkers Prev* **26**, 1114-1123, doi:10.1158/1055-9965.EPI-16-1012 (2017).
14 Chattopadhyay, P. K., Winters, A. F., Lomas, W. E., 3rd, Laino, A. S. & Woods, D. M. High-Parameter Single-Cell Analysis. *Annu Rev Anal Chem (Palo Alto Calif)* **12**, 411-430, doi:10.1146/annurev-anchem-061417-125927 (2019).
15 Woods, D. M. *et al.* Nivolumab and ipilimumab are associated with distinct immune landscape changes and response-associated immunophenotypes. *JCI Insight* **5**, doi:10.1172/jci.insight.137066 (2020).
16 Nettey, L., Giles, A. J. & Chattopadhyay, P. K. OMIP-050: A 28-color/30-parameter Fluorescence Flow Cytometry Panel to Enumerate and Characterize Cells Expressing a Wide Array of Immune Checkpoint Molecules. *Cytometry A* **93**, 1094-1096, doi:10.1002/cyto.a.23608 (2018).
17 Graze, P. R., Perlin, E. & Royston, I. In vitro lymphocyte dysfunction in Hodgkin's disease. *J Natl Cancer Inst* **56**, 239-243, doi:10.1093/jnci/56.2.239 (1976).
18 Case, D. C. *et al.* Comparison of multiple in vivo and in vitro parameters in untreated patients with Hodgkin's disease. *Cancer* **38**, 1807-1815, doi:10.1002/1097-0142(197610)38:4<1807::aid-cncr2820380458>3.0.co;2-y (1976).
19 Miller, D. G., Lizardo, J. G. & Snyderman, R. K. Homologous and heterologous skin transplantation in patients with lymphomatous disease. *J Natl Cancer Inst* **26**, 569-583 (1961).
20 Eltringham, J. R. & Kaplan, H. S. Impaired delayed-hypersensitivity responses in 154 patients with untreated Hodgkin's disease. *Natl Cancer Inst Monogr* **36**, 107-115 (1973).
21 Green, I. & Corso, P. F. A study of skin homografting in patients with lymphomas. *Blood* **14**, 235-245 (1959).
22 Skal, J. a. P., N. The delayed skin test response in Hodgkin’s disease and lymphosarcoma. *Cancer* **14**, 10 (1961).
23 Das, M., Zhu, C. & Kuchroo, V. K. Tim-3 and its role in regulating anti-tumor immunity. *Immunol Rev* **276**, 97-111, doi:10.1111/imr.12520 (2017).
24 Knee, D. A., Hewes, B. & Brogdon, J. L. Rationale for anti-GITR cancer immunotherapy. *Eur J Cancer* **67**, 1-10, doi:10.1016/j.ejca.2016.06.028 (2016).
25 Diefenbach, C. S. *et al.* Hodgkin Lymphoma: Current Status and Clinical Trial Recommendations. *J Natl Cancer Inst* **109**, doi:10.1093/jnci/djw249 (2017).

Figure 1

A)



B)



Combinatorial Phenotypes	Examples
Single marker	*CD8+*, *CD8-*, *CD45RO+*, *IFNG-*
Two markers	*CD8+CD45RO-*, *CD45RO+IFNG+*
Three markers	*CD8+CD45RO+IFNG+*, *CD8+IFNG-TNF+*
Four markers	*CD8+CD45RO-IFNG+TNF-*

Total number of phenoytpes = 3^x, where x = *number of markers*

C)

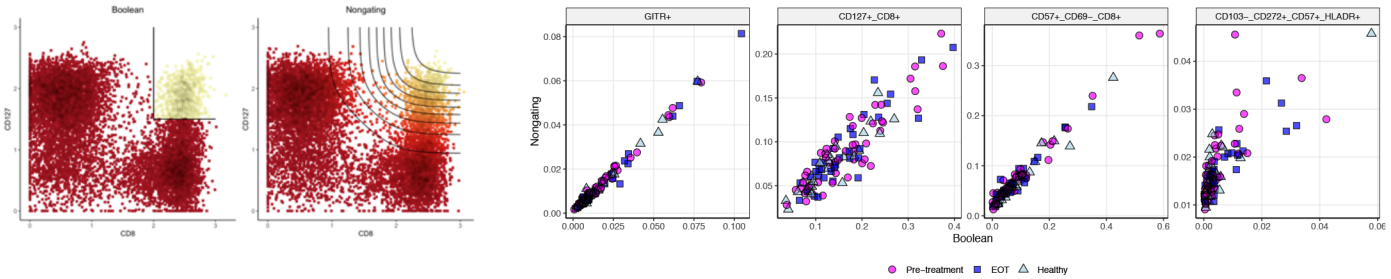


Figure 2

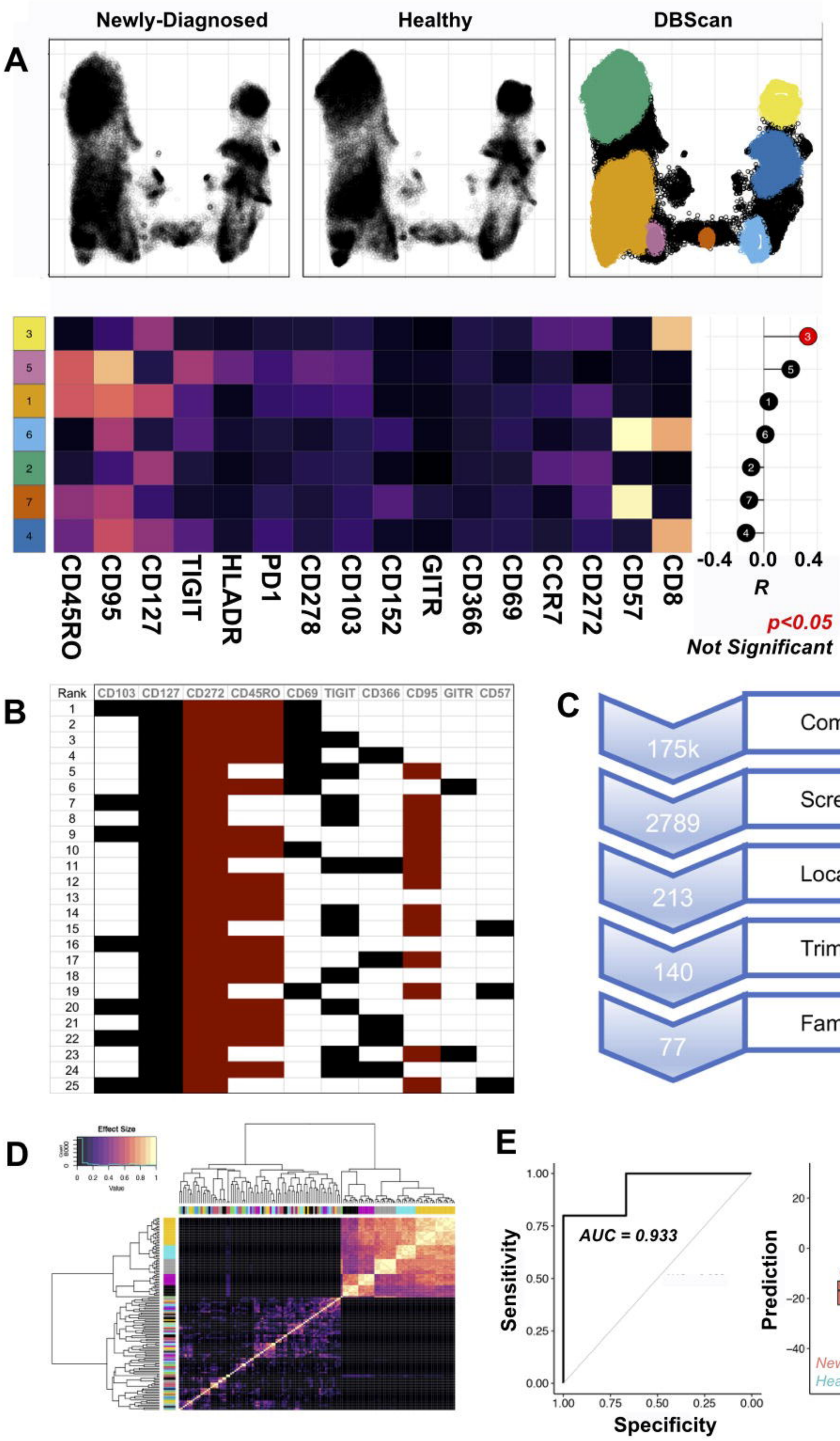

A
Newly-Diagnosed
Healthy
DBScan
3
5
1
6
2
7
4
CD45RO
CD95
CD127
TIGIT
HLADR
PD1
CD278
CD103
CD152
GITR
CD366
CD69
CCR7
CD272
CD57
CD8
-0.4
0.0
0.4
R
p<0.05
Not Significant
B
Rank
CD103
CD127
CD272
CD45RO
CD69
TIGIT
CD366
CD95
GITR
CD57
C
175k
Combinatorics
2789
Screening
213
Local maxima
140
Trimmed
77
Families
D
Effect Size
Value
E
AUC = 0.933
Sensitivity
Specificity
1.00
0.75
0.50
0.25
0.00
p<2E-16
p=0.024
Prediction
Newly-diagnosed
Healthy
Train
Test

Figure 3

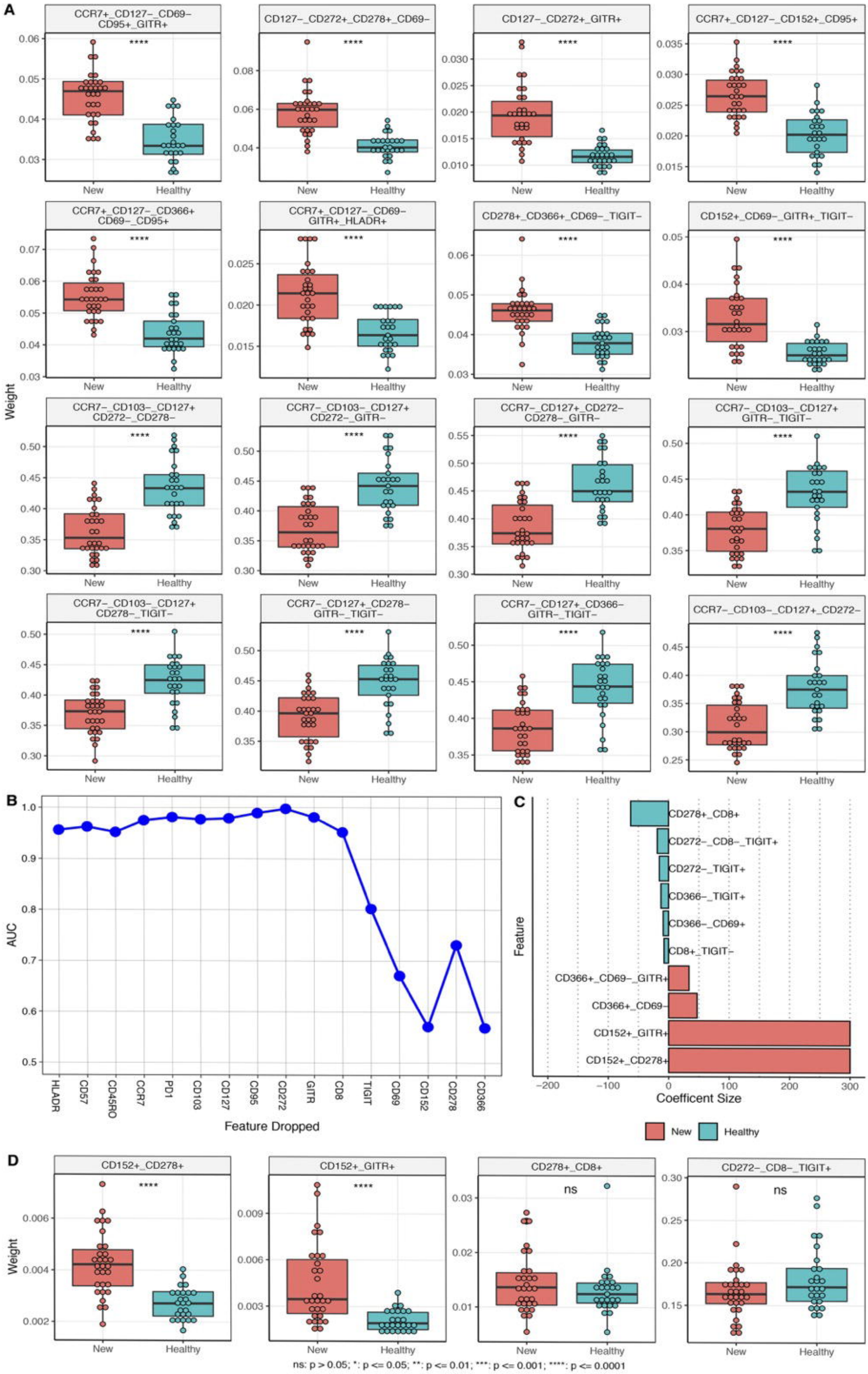
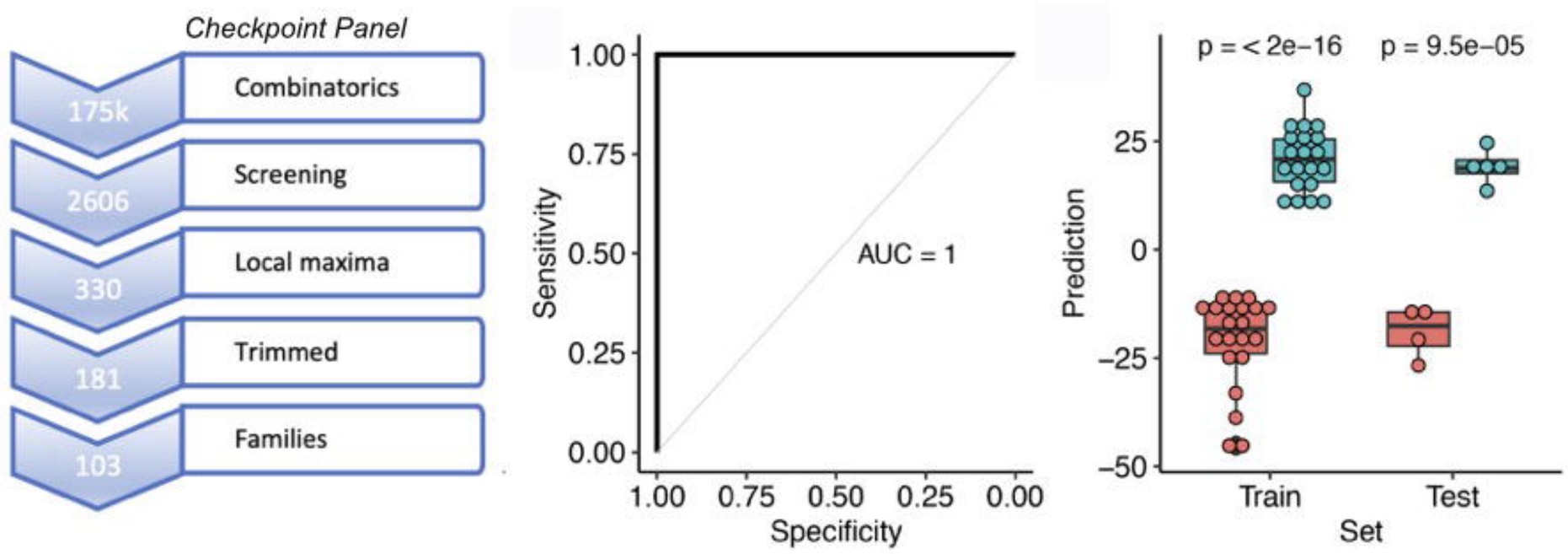
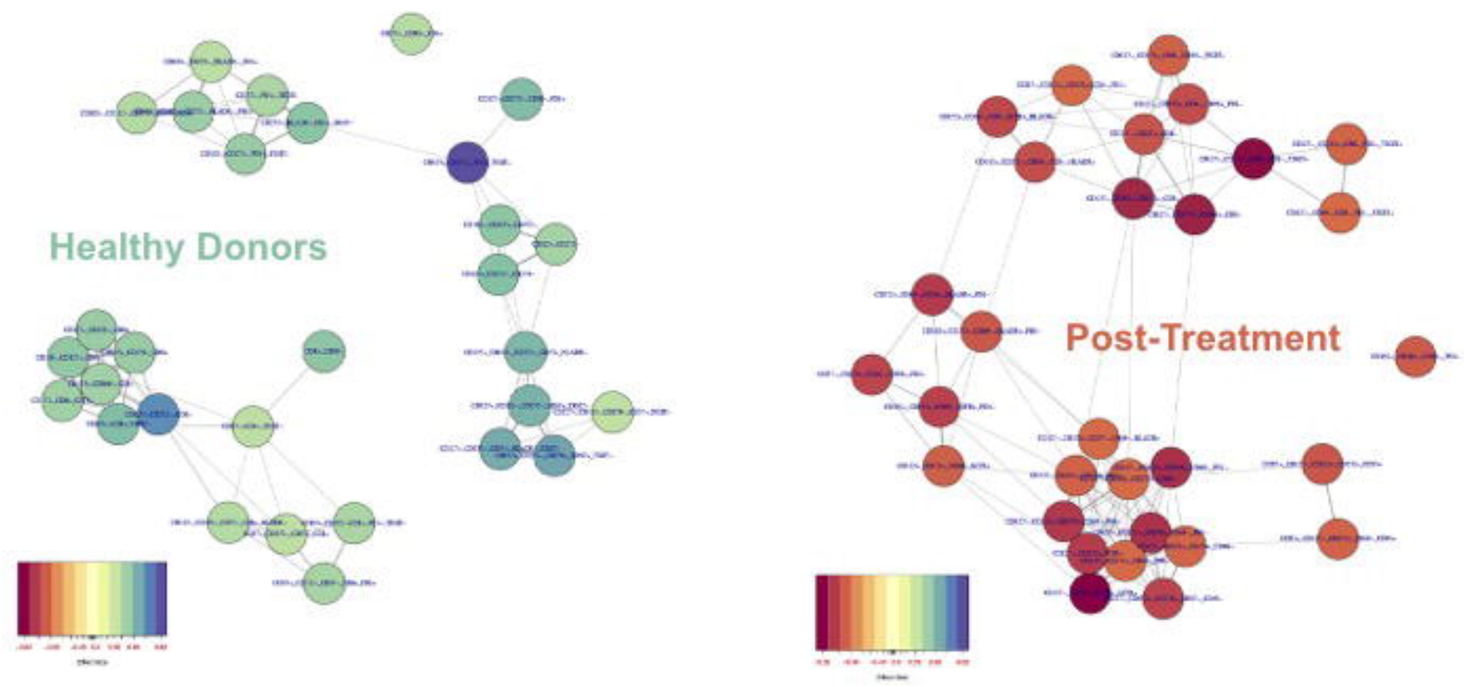

A
CCR7+_CD127−_CD69−_CD95+_GITR+
CD127−_CD272+_CD278+_CD69−
CD127−_CD272+_GITR+
CCR7+_CD127−_CD152+_CD95+
CCR7+_CD127−_CD366+_CD69−_CD95+
CCR7+_CD127−_CD69−_GITR+_HLADR+
CD278+_CD366+_CD69−_TIGIT−
CD152+_CD69−_GITR+_TIGIT−
CCR7−_CD103−_CD127+_CD272−_CD278−
CCR7−_CD103−_CD127+_CD272−_GITR−
CCR7−_CD127+_CD272−_CD278−_GITR−
CCR7−_CD103−_CD127+_GITR−_TIGIT−
CCR7−_CD103−_CD127+_CD278−_TIGIT−
CCR7−_CD127+_CD278−_GITR−_TIGIT−
CCR7−_CD127+_CD366−_GITR−_TIGIT−
CCR7−_CD103−_CD127+_CD272−
Weight
New
Healthy
****
B
AUC
Feature Dropped
HLADR
CD57
CD45RO
CCR7
PD1
CD103
CD127
CD95
CD272
GITR
CD8
TIGIT
CD69
CD152
CD278
CD366
C
Feature
CD278+_CD8+
CD272−_CD8−_TIGIT+
CD272−_TIGIT+
CD366−_TIGIT+
CD366−_CD69+
CD8+_TIGIT−
CD366+_CD69−_GITR+
CD366+_CD69−
CD152+_GITR+
CD152+_CD278+
Coefficent Size
New
Healthy
D
CD152+_CD278+
CD152+_GITR+
CD278+_CD8+
CD272−_CD8−_TIGIT+
ns
ns: p > 0.05; *: p <= 0.05; **: p <= 0.01; ***: p <= 0.001; ****: p <= 0.0001

Figure 4

A


Checkpoint Panel
175k
Combinatorics
2606
Screening
330
Local maxima
181
Trimmed
103
Families
Sensitivity
Specificity
AUC = 1
Prediction
p = < 2e−16
p = 9.5e−05
Train
Test
Set

B


Healthy Donors
Post-Treatment

C

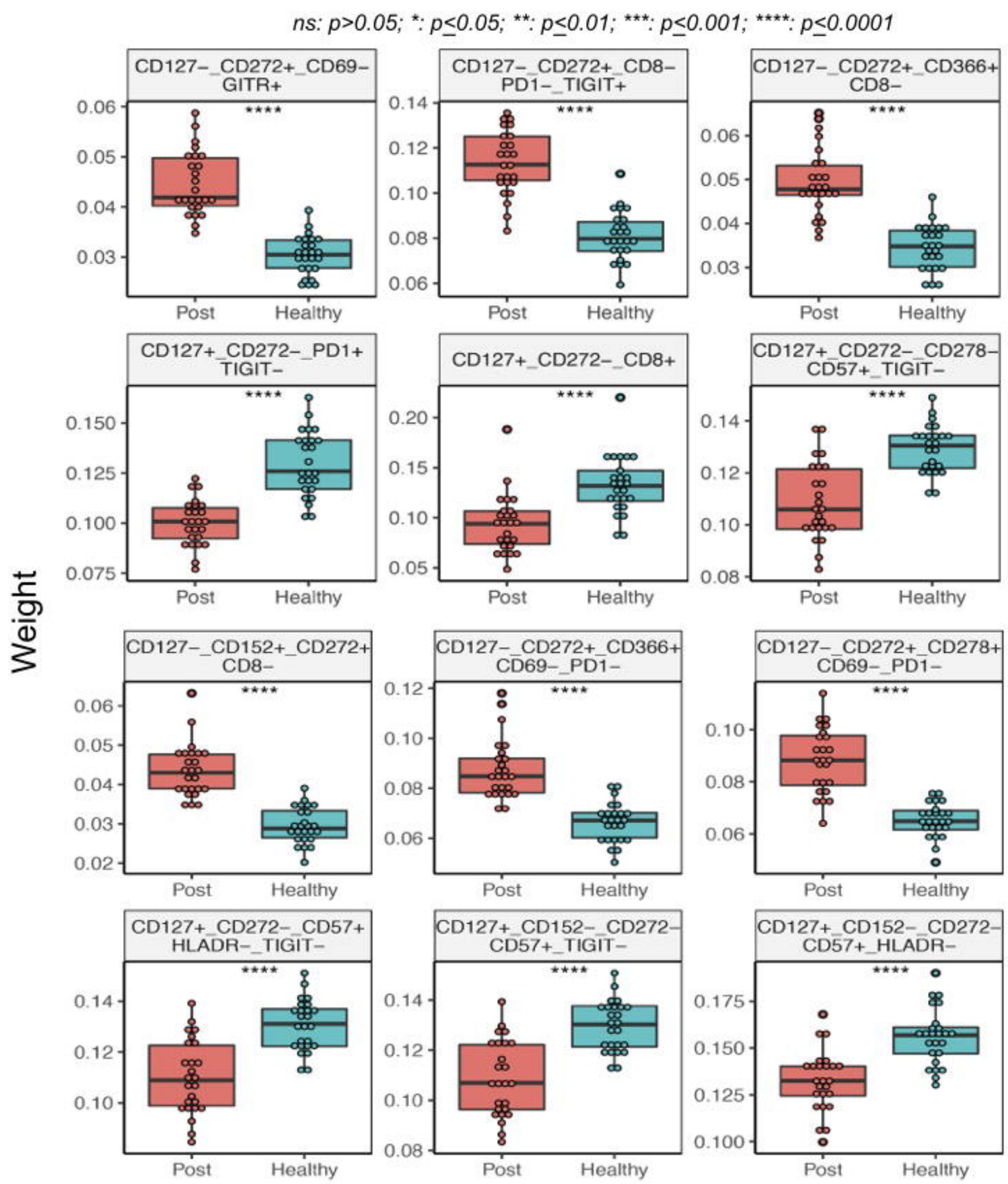
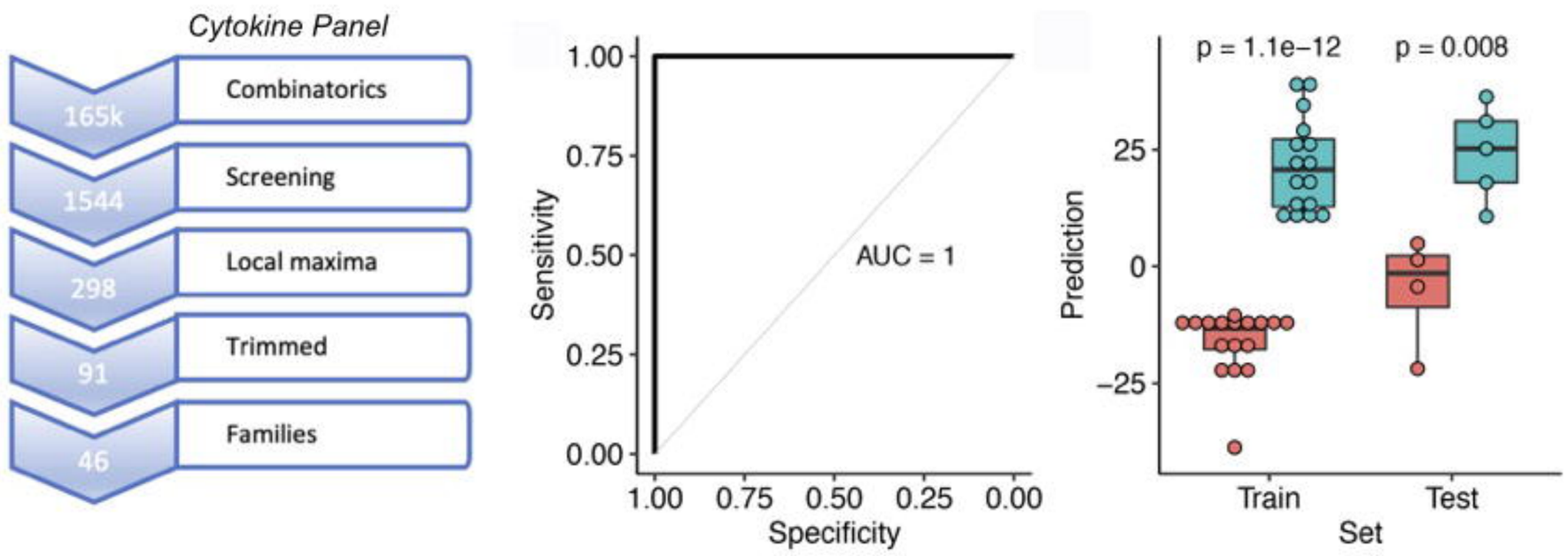
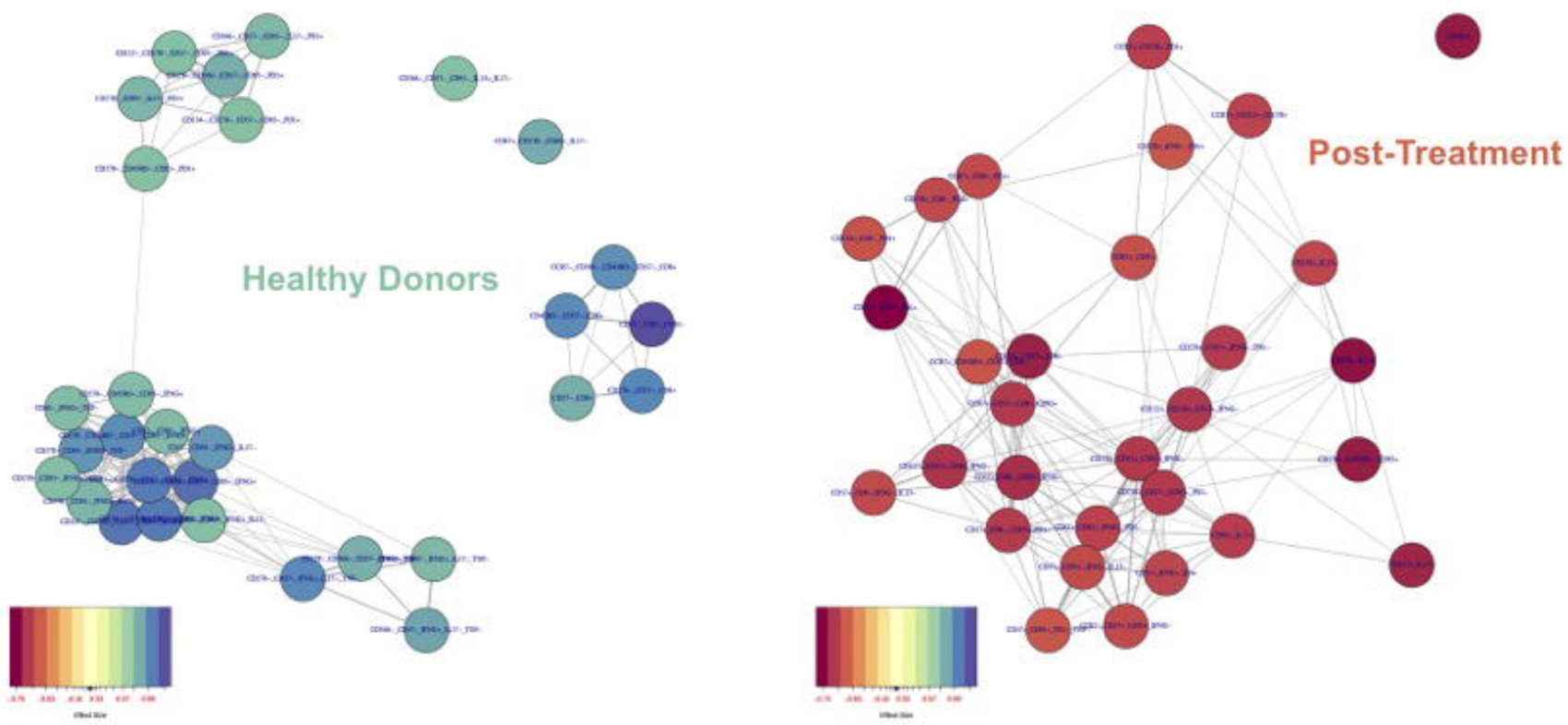

ns: p>0.05; *: p≤0.05; **: p≤0.01; ***: p≤0.001; ****: p≤0.0001
Weight
CD127−_CD272+_CD69−_GITR+
CD127−_CD272+_CD8−_PD1−_TIGIT+
CD127−_CD272+_CD366+_CD8−
CD127+_CD272−_PD1+_TIGIT−
CD127+_CD272−_CD8+
CD127+_CD272−_CD278−_CD57+_TIGIT−
CD127−_CD152+_CD272+_CD8−
CD127−_CD272+_CD366+_CD69−_PD1−
CD127−_CD272+_CD278+_CD69−_PD1−
CD127+_CD272−_CD57+_HLADR−_TIGIT−
CD127+_CD152−_CD272−_CD57+_TIGIT−
CD127+_CD152−_CD272−_CD57+_HLADR−
****
Post
Healthy

Figure 5

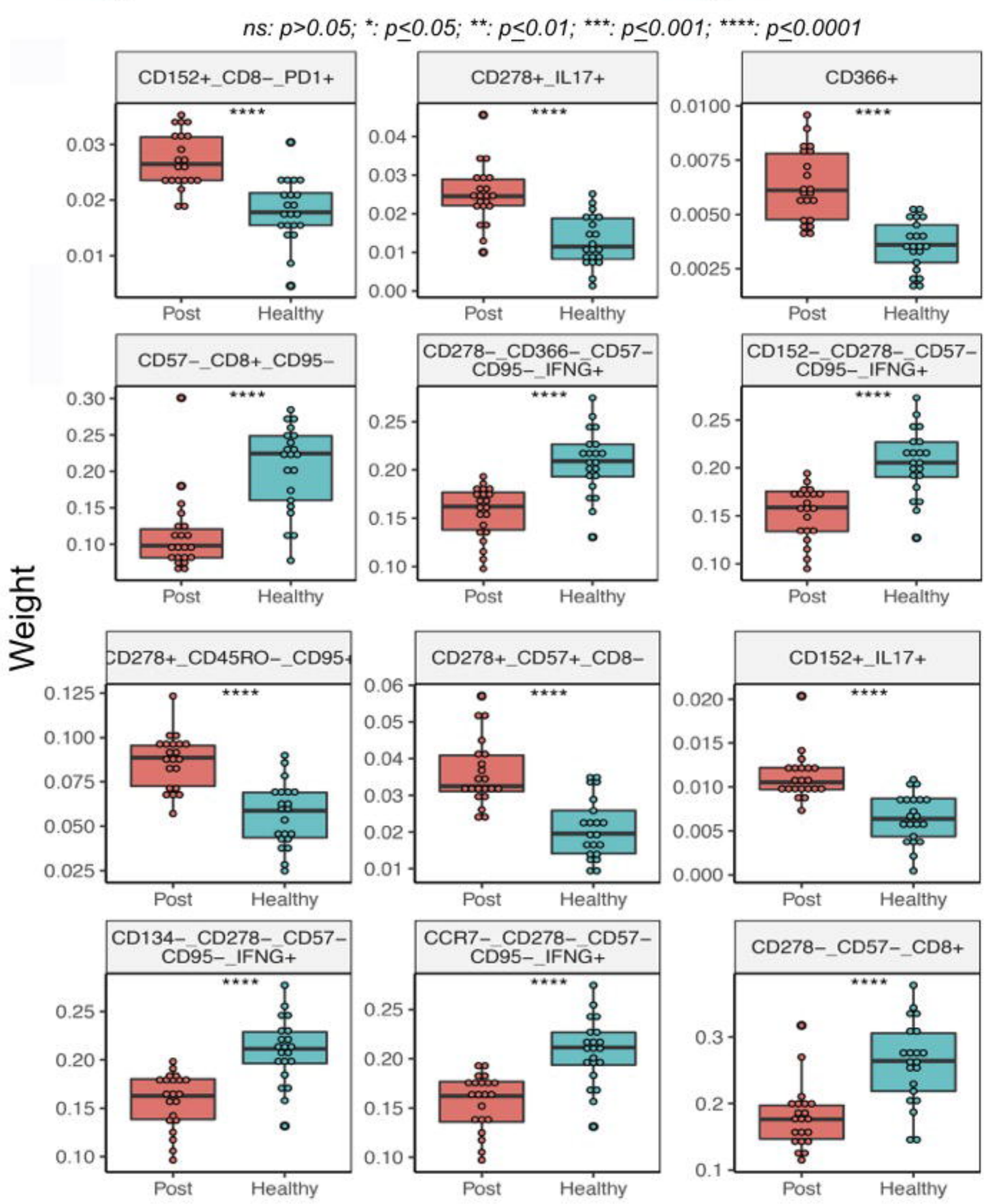
A


Cytokine Panel
165k
Combinatorics
1544
Screening
298
Local maxima
91
Trimmed
46
Families
Sensitivity
Specificity
AUC = 1
1.00
0.75
0.50
0.25
0.00
p = 1.1e−12
p = 0.008
Prediction
25
0
−25
Train
Test
Set

B


Healthy Donors
Post-Treatment

C


ns: p>0.05; *: p≤0.05; **: p≤0.01; ***: p≤0.001; ****: p≤0.0001
Weight
CD152+_CD8−_PD1+
CD278+_IL17+
CD366+
CD57−_CD8+_CD95−
CD278−_CD366−_CD57−_CD95−_IFNG+
CD152−_CD278−_CD57−_CD95−_IFNG+
CD278+_CD45RO−_CD95+
CD278+_CD57+_CD8−
CD152+_IL17+
CD134−_CD278−_CD57−_CD95−_IFNG+
CCR7−_CD278−_CD57−_CD95−_IFNG+
CD278−_CD57−_CD8+
****
Post
Healthy

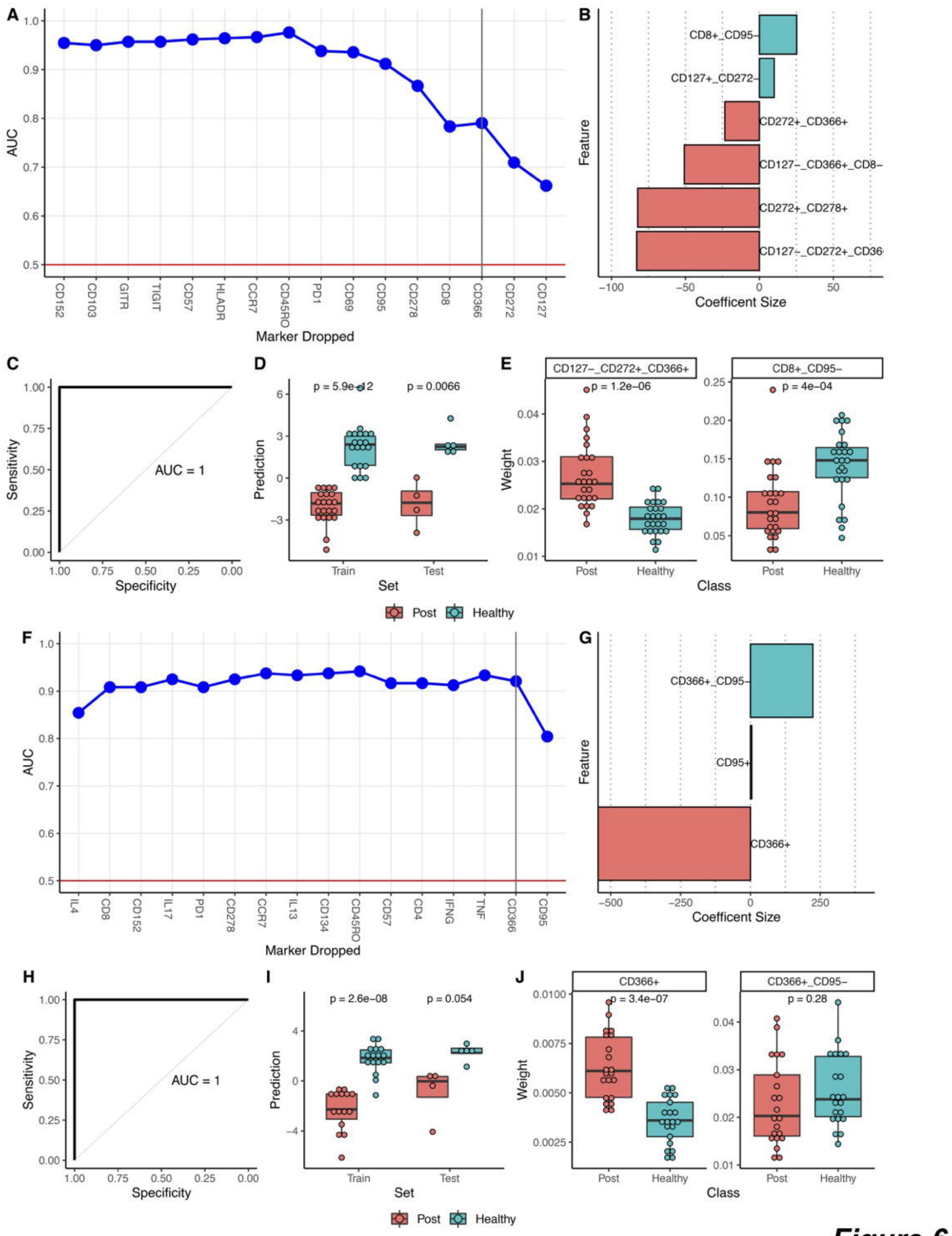
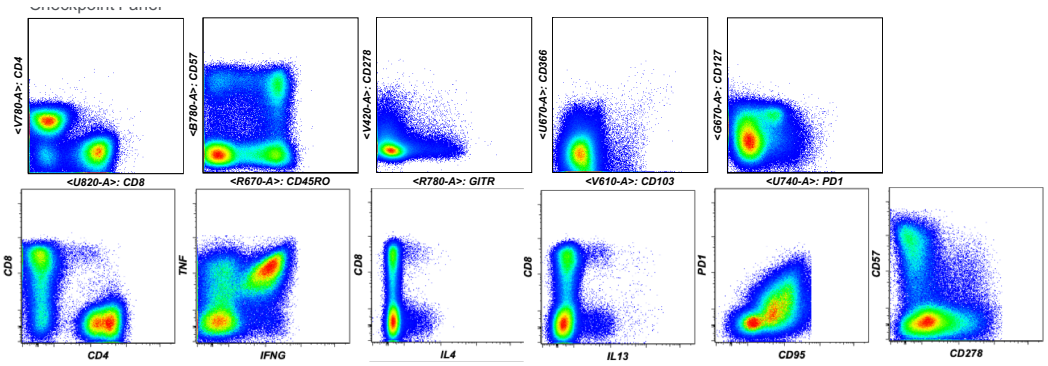

A
AUC
Marker Dropped
CD152
CD103
GITR
TIGIT
CD57
HLADR
CCR7
CD45RO
PD1
CD69
CD95
CD278
CD8
CD366
CD272
CD127
B
Feature
CD8+_CD95−
CD127+_CD272−
CD272+_CD366+
CD127−_CD366+_CD8−
CD272+_CD278+
CD127−_CD272+_CD36
Coefficent Size
C
Sensitivity
Specificity
AUC = 1
D
p = 5.9e−12
p = 0.0066
Prediction
Train
Test
Set
E
CD127−_CD272+_CD366+
p = 1.2e−06
CD8+_CD95−
p = 4e−04
Weight
Post
Healthy
Class
Post
Healthy
F
AUC
IL4
CD8
CD152
IL17
PD1
CD278
CCR7
IL13
CD134
CD45RO
CD57
CD4
IFNG
TNF
CD366
CD95
Marker Dropped
G
Feature
CD366+_CD95−
CD95+
CD366+
Coefficent Size
H
Sensitivity
Specificity
AUC = 1
I
p = 2.6e−08
p = 0.054
Prediction
Train
Test
Set
Post
Healthy
J
CD366+
p = 3.4e−07
CD366+_CD95−
p = 0.28
Weight
Post
Healthy
Class

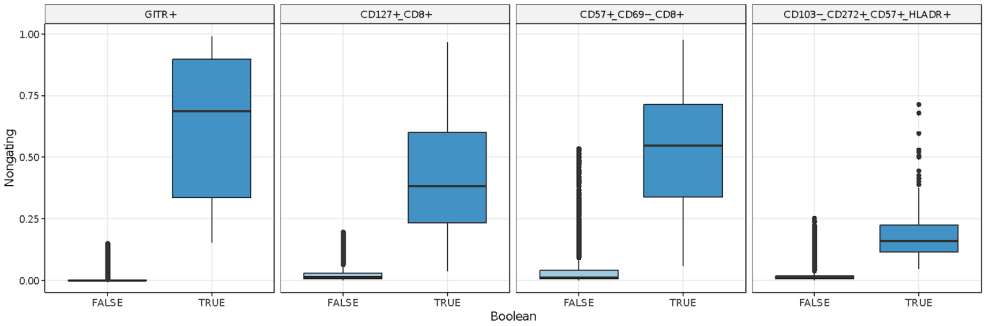
Figure 6

Supplemental Figure 1

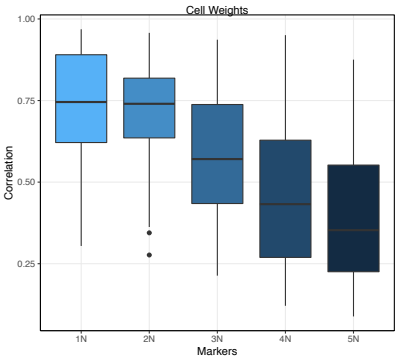
A



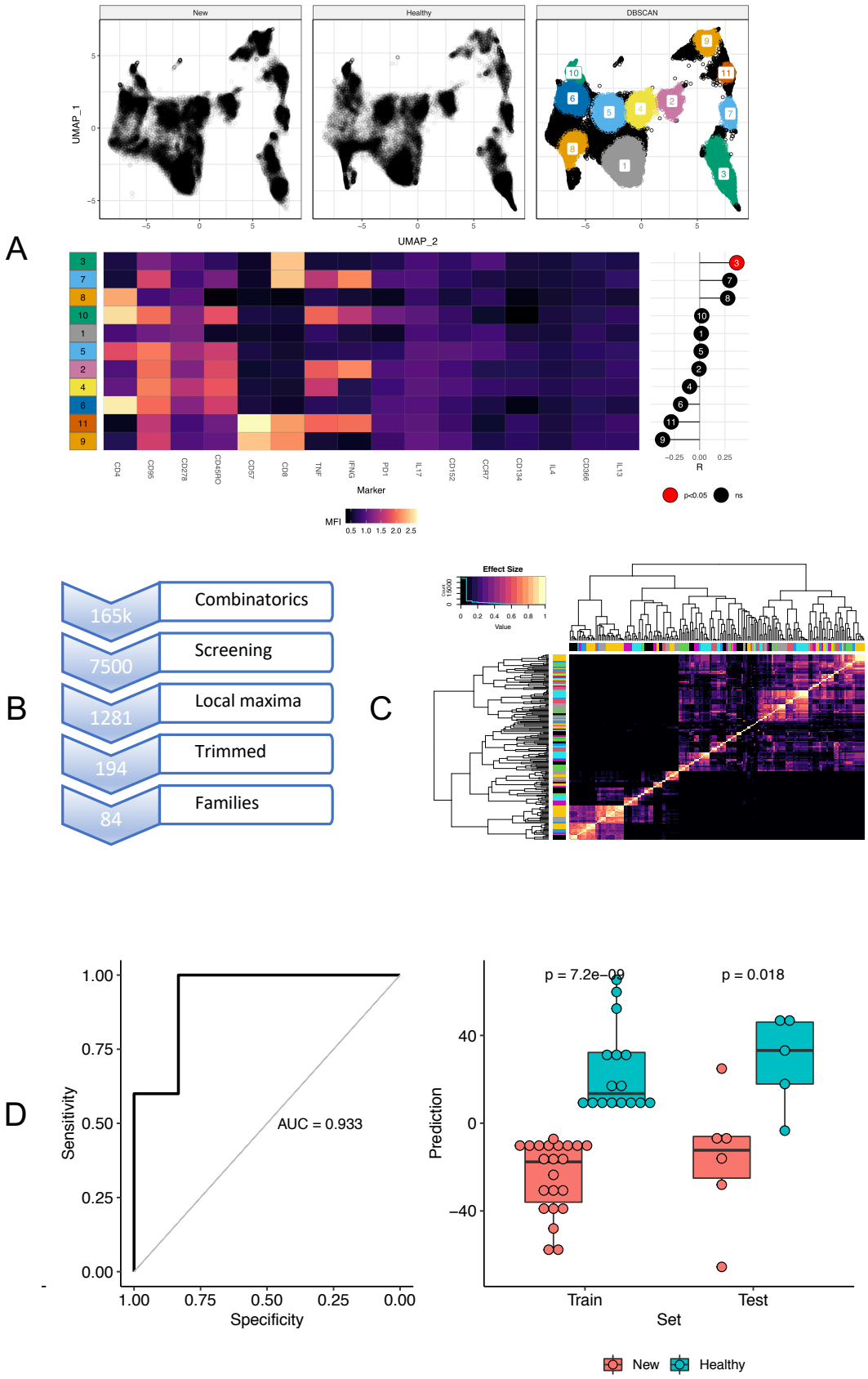
B



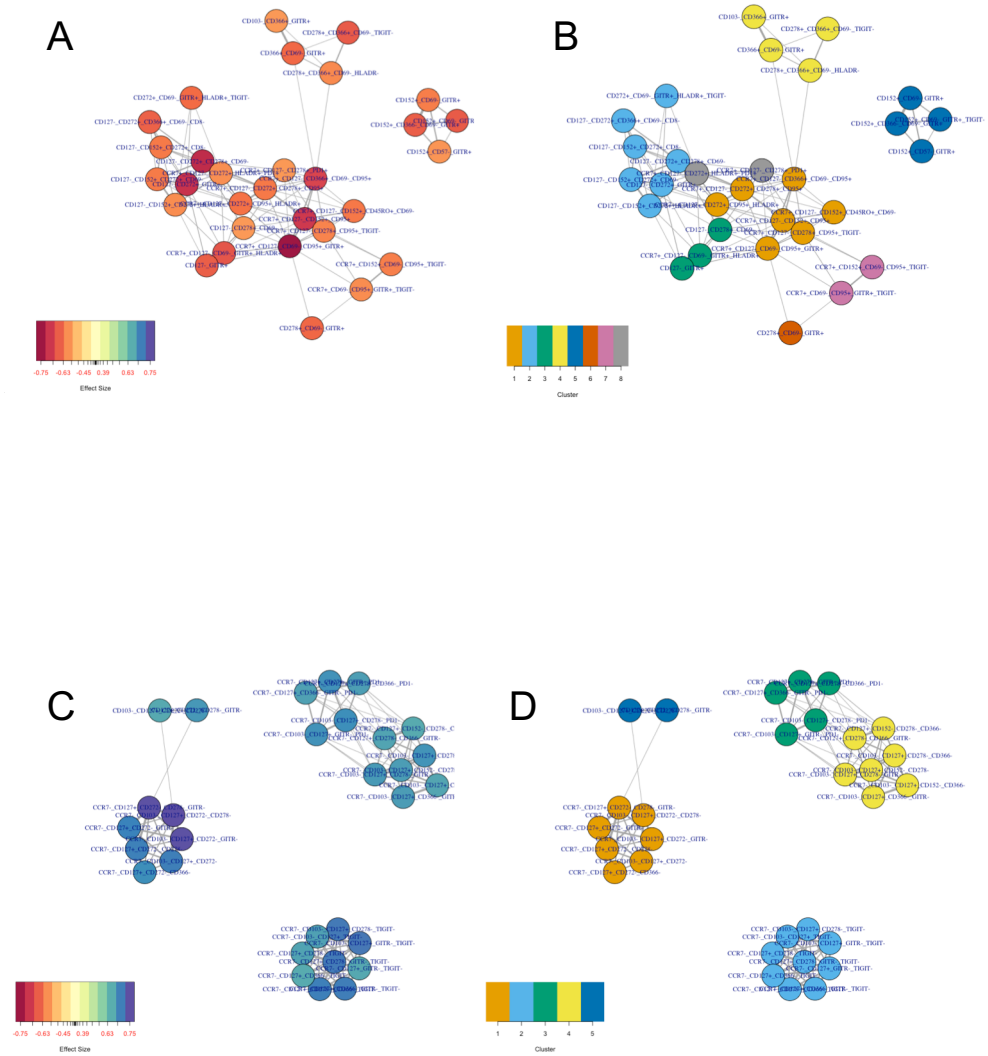
C



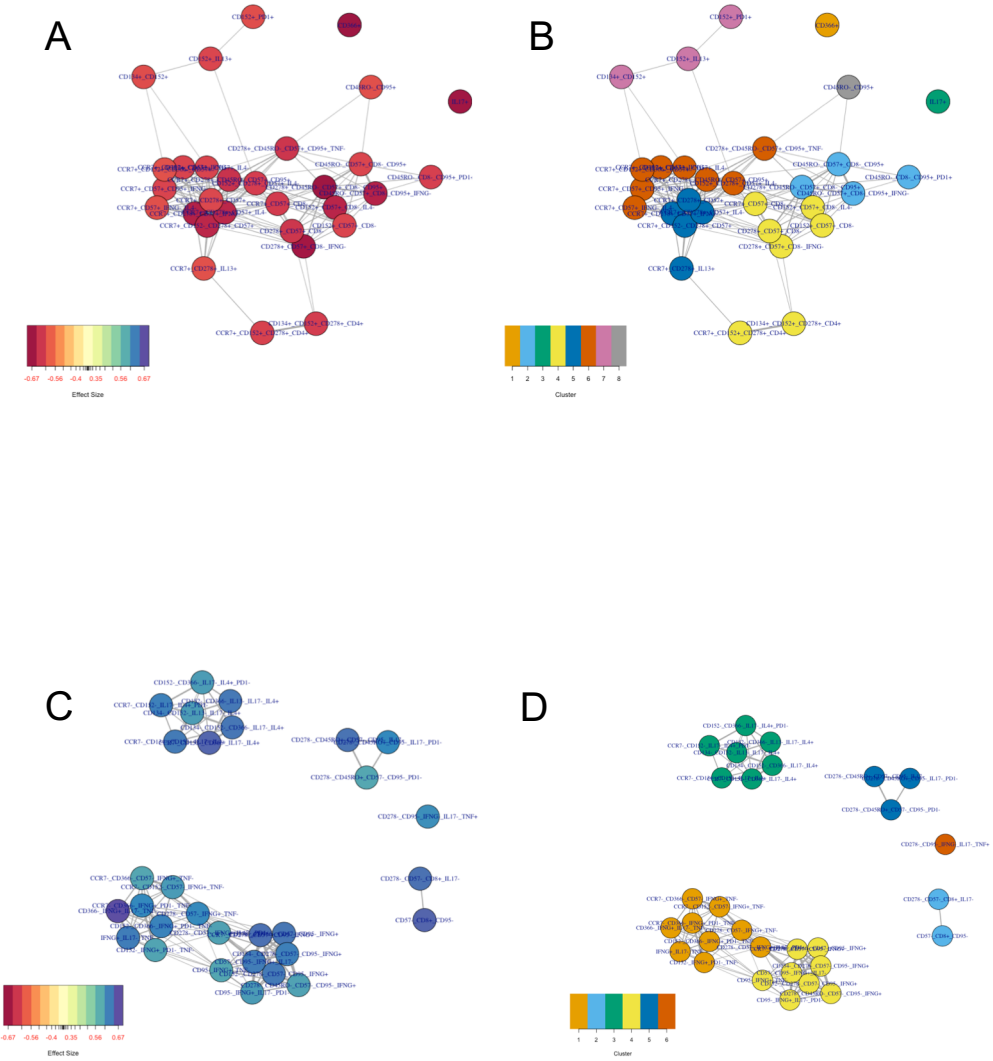
Supplemental Figure 2



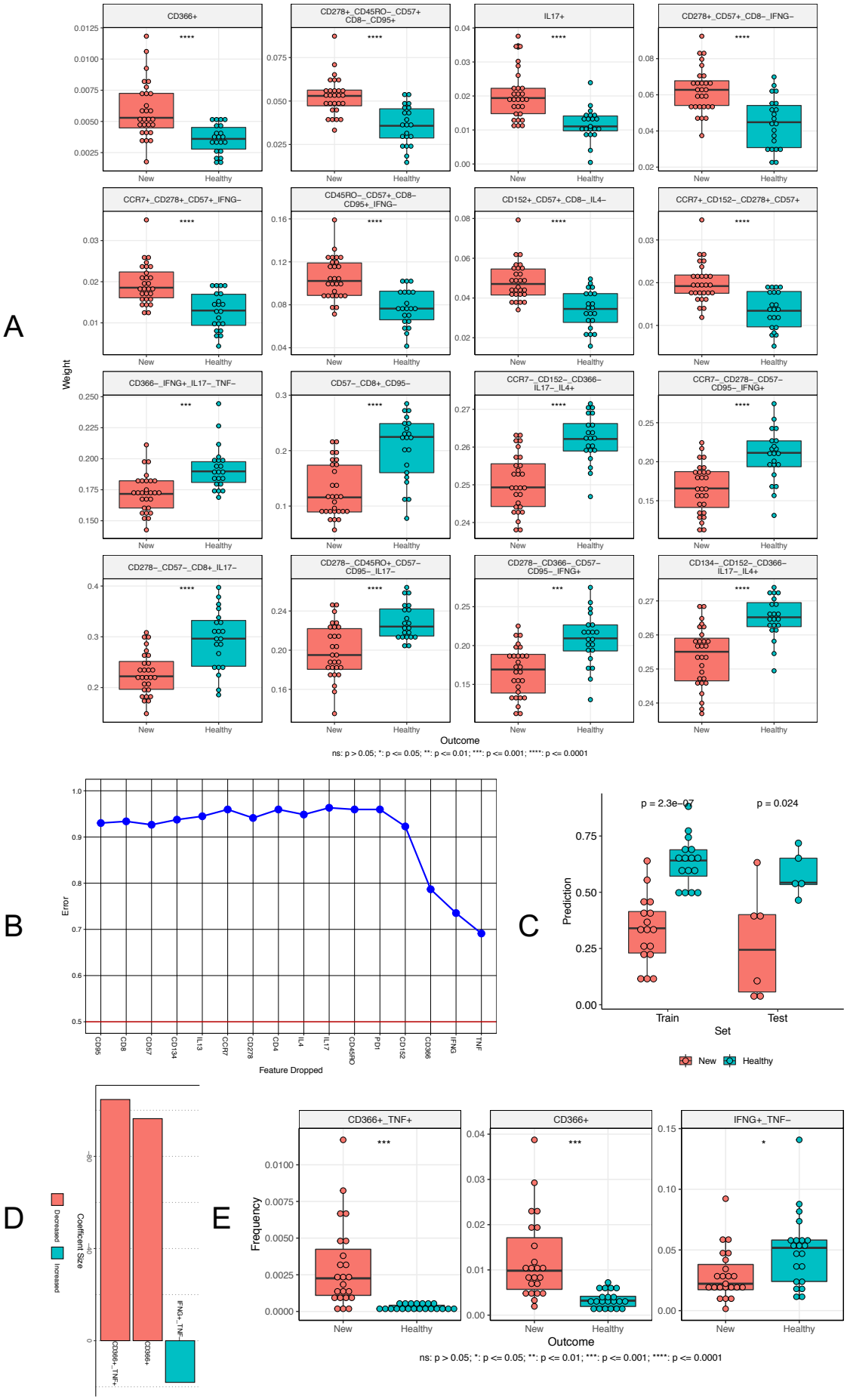
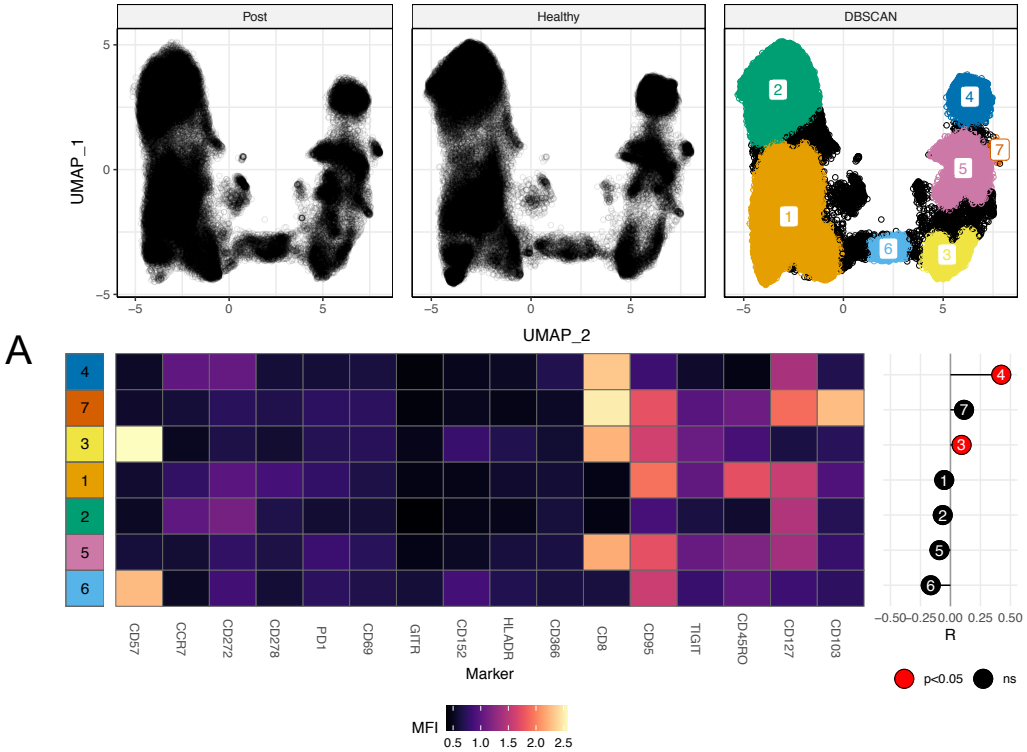
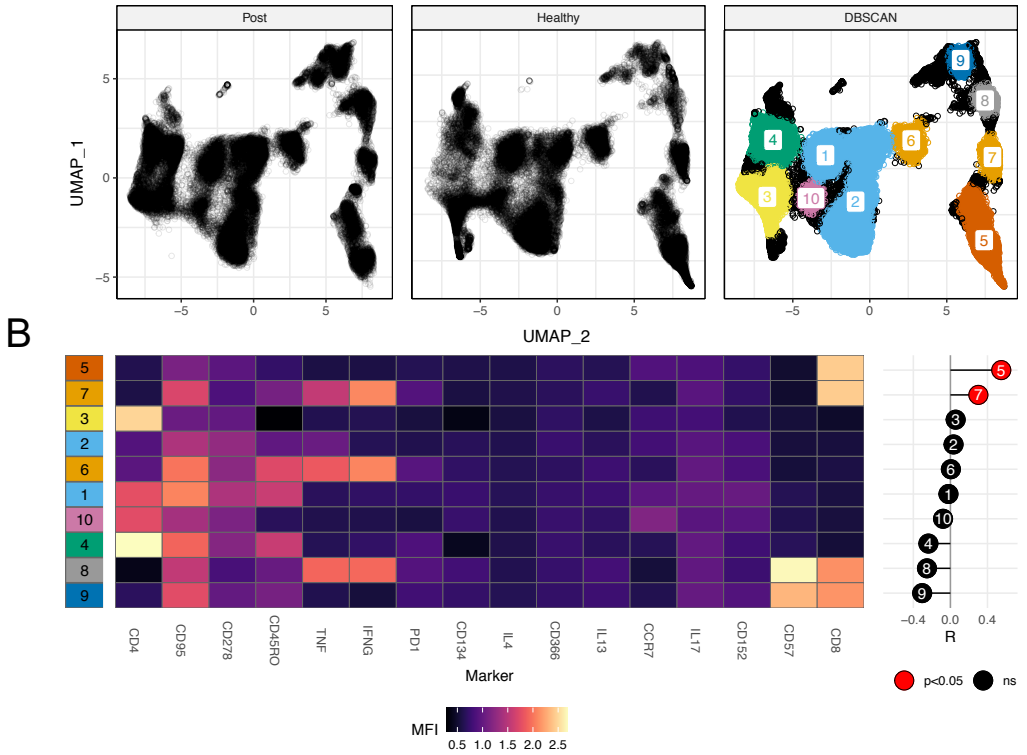
Supplemental Figure 3



Supplemental Figure 4



Supplemental Figure 5



Supplemental Figure 6