–SUPPLEMENTARY INFORMATION–

Restricting datasets to classifiable samples augments discovery of immune disease markers

Gunther Glehr[1], Paloma Riquelme[1], Katharina Kronenberg[1], Robert Lohmayer[2], Víctor J. López-Madrona[3], Michael Kapinsky[4], Hans J. Schlitt[1], Edward K. Geissler[1], Rainer Spang[5], Sebastian Haferkamp[6], and James A. Hutchinson[1,*]

[1]Department of Surgery, University Hospital Regensburg, Regensburg, Germany
[2]Algorithmic Bioinformatics Research Group, Leibniz Institute for Immunotherapy, Regensburg, Germany
[3]Aix Marseille Univ, INSERM, Institut de Neurosciences des Systèmes, Marseille, France
[4]Beckman Coulter Life Sciences GmbH, Krefeld, Germany
[5]Department of Statistical Bioinformatics, University of Regensburg, Regensburg, Germany
[6]Department of Dermatology, University Hospital Regensburg, Regensburg, Germany
[*]Corresponding author: James A. Hutchinson, james.hutchinson@ukr.de

Supplementary Notes

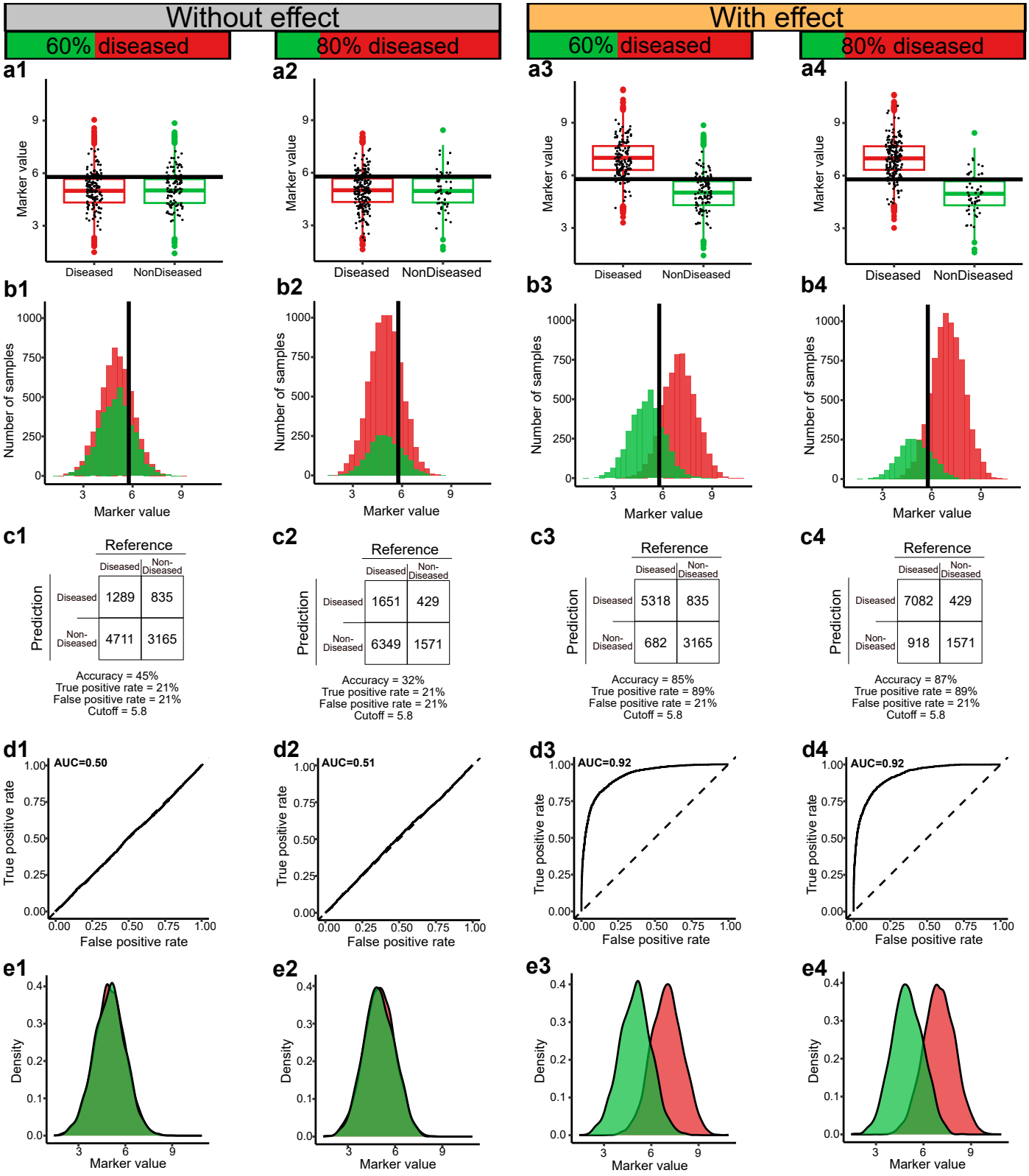
Supplementary Figures

Supplementary Tables

Supplementary Videos

Supplementary Websites

Supplementary Note 1: Class sizes and discriminatory value



Supplementary Figure 1 *Class sizes do not determine the informative value of disease markers.*

In this supplementary note, we explain the relationship between the discriminatory value of disease markers, class sizes and correct classification rates. Most readers intuitively appreciate that a test's accuracy (also called the correct classification rate) depends upon the relative numbers of cases in each class. However, it may be confusing that a marker's discriminatory value is independent of class sizes. Figure 1 explains this critical distinction through four illustrated examples, arranged in columns (1 - 4) that each consists of 5 panels (a - e).

The vertically arranged panels depict the same datasets treated in five different ways:

(a) Boxplot of the marker values for both diseased (red) and non-diseased (green) samples. 250 randomly selected samples represented as individual points. The boxplot provides a visual summary of the distribution of the marker values. An arbitrary discriminatory cut-off is displayed at a marker value of 5.8.

(b) Histogram of marker values for diseased (red) and non-diseased (green) samples. The histogram illustrates the distribution of cases with respect to marker values in each class. The same arbitrary discriminatory cut-off is displayed at a marker value of 5.8.

(c) Confusion matrix summarizing the predicted class according to the arbitrary marker value cut-off value of 5.8. This table shows the number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FP). From these numbers, we can calculate the true positive rate (TPR), false positive rate (FPR) and accuracy. Hence, the confusion matrix helps to evaluate the performance of a marker in predicting disease status for one specific cut-off value.

(d) Receiver operating characteristic (ROC) curves are graphical representations of the performance of a marker in distinguishing between diseased and non-diseased samples. ROC curves plot the true positive rate (sensitivity) against the false positive rate (1-specificity) calculated for all marker values in the dataset. As described in Figure 1, area under the ROC (AUC) is a measure of a marker's discriminatory power.

(e) Densities for diseased (red) and non-diseased samples. Imagine as histogram divided by the total number of samples in the respective class.

Let us first consider columns 1 and 2, which represent a marker with no discriminatory value:

(a) Regardless of the number of diseased (red: column 1, n=6000; column 2, n=8000) or non-diseased (green: column 1, n=4000; column 2, n=2000) samples in each class, the boxplots shown in (a1) and (a2) are almost indistinguishable.

(b) In contrast, the histograms in (b1) and (b2) clearly show more diseased samples in column 2.

(c) By definition, no marker value is useful for discriminating between the two classes. Our arbitrary choice of 5.8 as a cut-off leads to confusion tables (c1) and (c2). The accuracy of this test decreases from (c1; 45%) to (c2; 32%) because more samples happen to fall below the cut-off. Importantly, TPR (21%) and FPR (21%) are identical irrespective of class size.

(d) The ROC curves in (d1) and (d2) plot TPR against FPR for these samples. Because TPR and FPR are not affected by class size, the ROC curves and their area under the curve are essentially identical (d1: AUC = 0.50; d2: AUC = 0.51). Importantly, our interpretation of these curves as showing a marker with no discriminatory value is not affected by the different class sizes in (d1) and (d2).

(e) The density plots shown in (e1) and (e2) help us to understand why our hypothetical marker has no discriminatory power. By normalizing the class sizes, we see the marker distributions in the diseased (red) and non-diseased (green) classes are essentially identically shaped and completely overlapping. Hence, it is correct to say that ROC curves are based upon densities, not absolute numbers of cases.

Let us now consider columns 3 and 4, which represent a marker with a high discriminatory value:

(a) Regardless of the number of diseased (red: column 3, n=6000; column 4, n=8000) or non-diseased (green: column 3, n=4000; column 4, n=2000) samples in each class, the boxplots shown in (a1) and (a2) are almost indistinguishable.

(b) In contrast, the histograms in (b3) and (b4) clearly show more diseased samples in column 4.

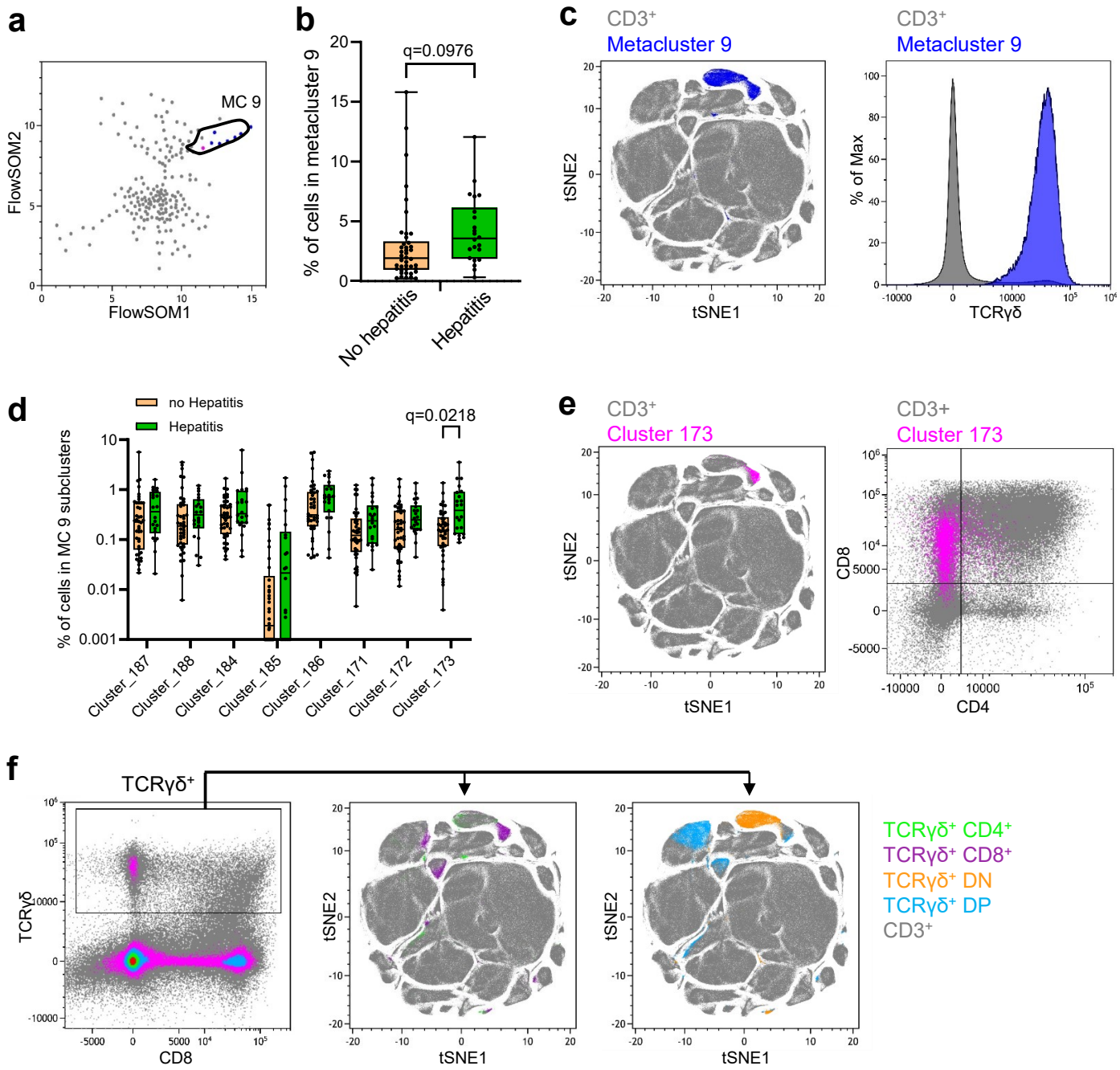
(c) Our arbitrary choice of 5.8 as a cut-off leads to confusion tables (c3) and (c4). The accuracy of this test increases slightly from (c3; 85%) to (c4; 87%) owing to the relative differences in class sizes. Importantly, TPR (89%) and FPR (21%) are identical irrespective of class size.

(d) The ROC curves in (d3) and (d4) plot TPR against FPR for these samples. Because TPR and FPR are not affected by class size, the ROC curves are essentially identical (d3: AUC = 0.92; d4: AUC = 0.92). Importantly, our interpretation of these curves as showing a marker with high discriminatory value is not affected by the different class sizes in (d3) and (d4).

(e) The density plots shown in (e1) and (e2) help us to understand why our hypothetical marker has high discriminatory power. By normalizing the class sizes, we see the marker distributions in the diseased (red) and non-diseased (green) classes are essentially identically shaped but only have a small overlap.

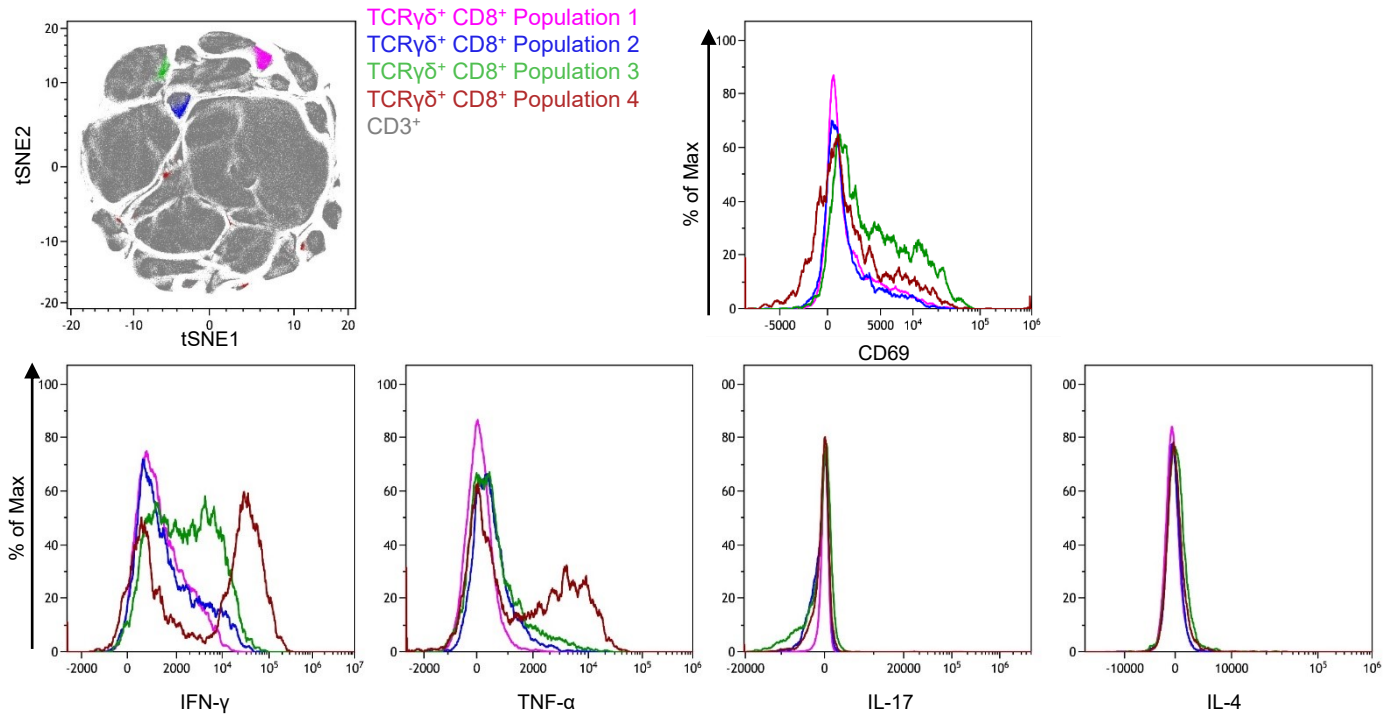
In summary, class sizes may affect the accuracy of a test, but not its TPR or FPR. Hence, the discriminatory value of a marker, which is a function of TPR and FPR, is not determined by class sizes. Densities are a class size-independent way of visualising the expression of a disease marker within a patient class. Because densities are directly relevant to a marker's discriminatory value, we present plots of densities throughout this article.

Supplementary Figure 2



Supplementary Figure 2 *part 1/2. FlowSOM clustering results in TCR$\gamma\delta^+$ cells as a marker for hepatitis.* **(a)** PBMC from 64 patients were stimulated for 4 hours and stained for cell surface antigen and intracellular cytokines. FlowSOM analysis was performed on CD3$^+$ cells with 9 metaclusters and 196 clusters. Metacluster 9 colored blue, cluster 173 colored pink (n=42 no hepatitis, n=22 hepatitis). **(b)** There was a higher abundance of cells in metacluster 9 in patients that subsequently developed hepatitis. **(c)** Metacluster 9 corresponded to a population of TCR$\gamma\delta^+$ cells. **(d)** The proportion of cells in cluster 173 was higher in hepatitis patients. **(e)** Cluster 173 included mostly CD8$^+$ TCR$\gamma\delta^+$ cells. **(f)** TCR$\gamma\delta^+$ cells are distributed in different areas of the viSNE plot; metacluster 9 was mainly formed by CD8$^+$ and double negative (DN) T cells.

g



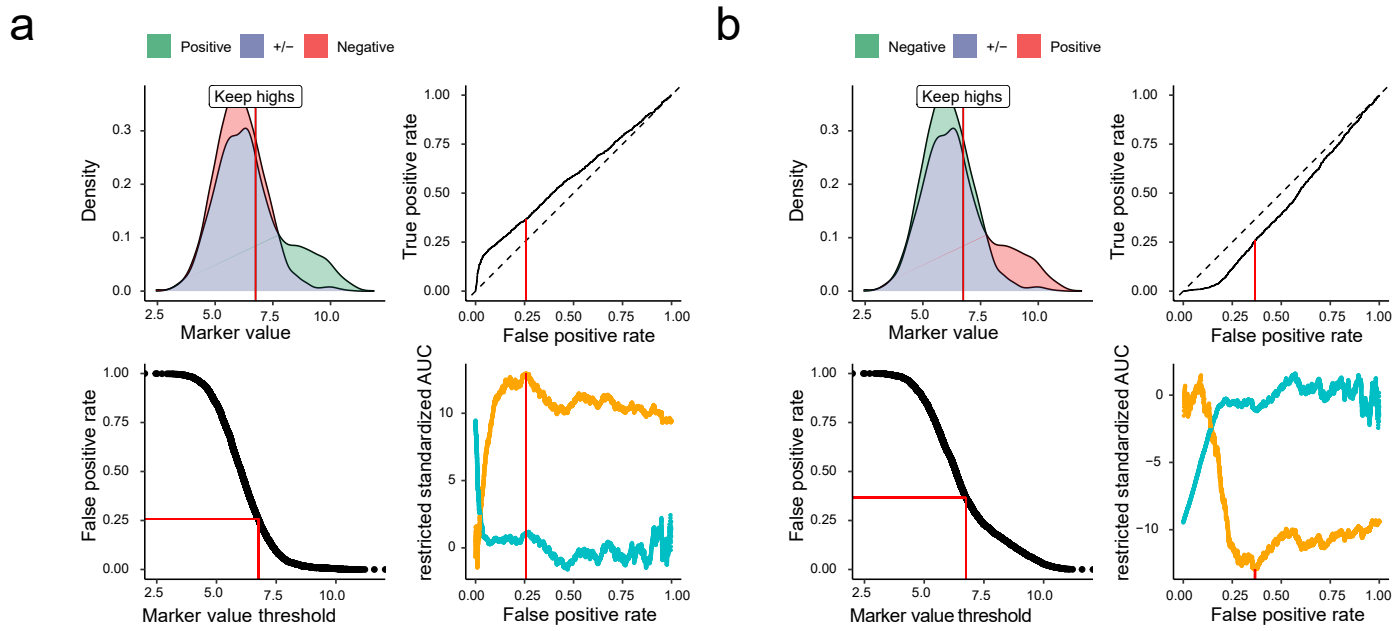
h

488 nm	561 nm				638 nm			405 nm				808 nm
525/40	585/42	610/20	710/50	763/43	660/10	712/25	763/43	450/45	525/40	610/20	763/43	885/40
AF488	**PE**	**AF594**	**PC5.5**	**PC7**	**AF647**	**AF700**	**AFire750**	**BV421**	**BV510**	BV605	**BV785**	Viability

Antigen	TCR vα7.2	IL-17	CD8	TCRγδ	TCR-Vα24	IL-4	CD4	CD3	IFNγ	TNFα	CD69	CD56	Live-Dead
Clone	3C10	BL168	RPA-T8	IMMU510	6B11	MP4-25D2	SK	UCHT1	B27	MAb11	FN50	5.1H11	
Isotype	mIgG1	mIgG1	mIgG1	mIgG1	mIgG1	rIgG1	mIgG1	mIgG1	mIgG1	mIgG1	mIgG1	mIgG1	
Fluorochrome	FITC	PE	AF594	PC5.5	PC7	APC	AF700	APC-A750	BV421	BV510	BV605	BV785	ViaKr-808
Vol. (µl)	3	2	3	6	3	1	3	5	2	2	5	5	2
Supplier	Biolegend	BioLegend	BioLegend	BC	BioLegend	BioLegend	BioLegend	BC	BioLegend	BioLegend	BioLegend	BioLegend	BC
Cat.#	351703	512306	301056	A99021	342911	500812	344621	A94680	506538	502950	310937	362549	C36628

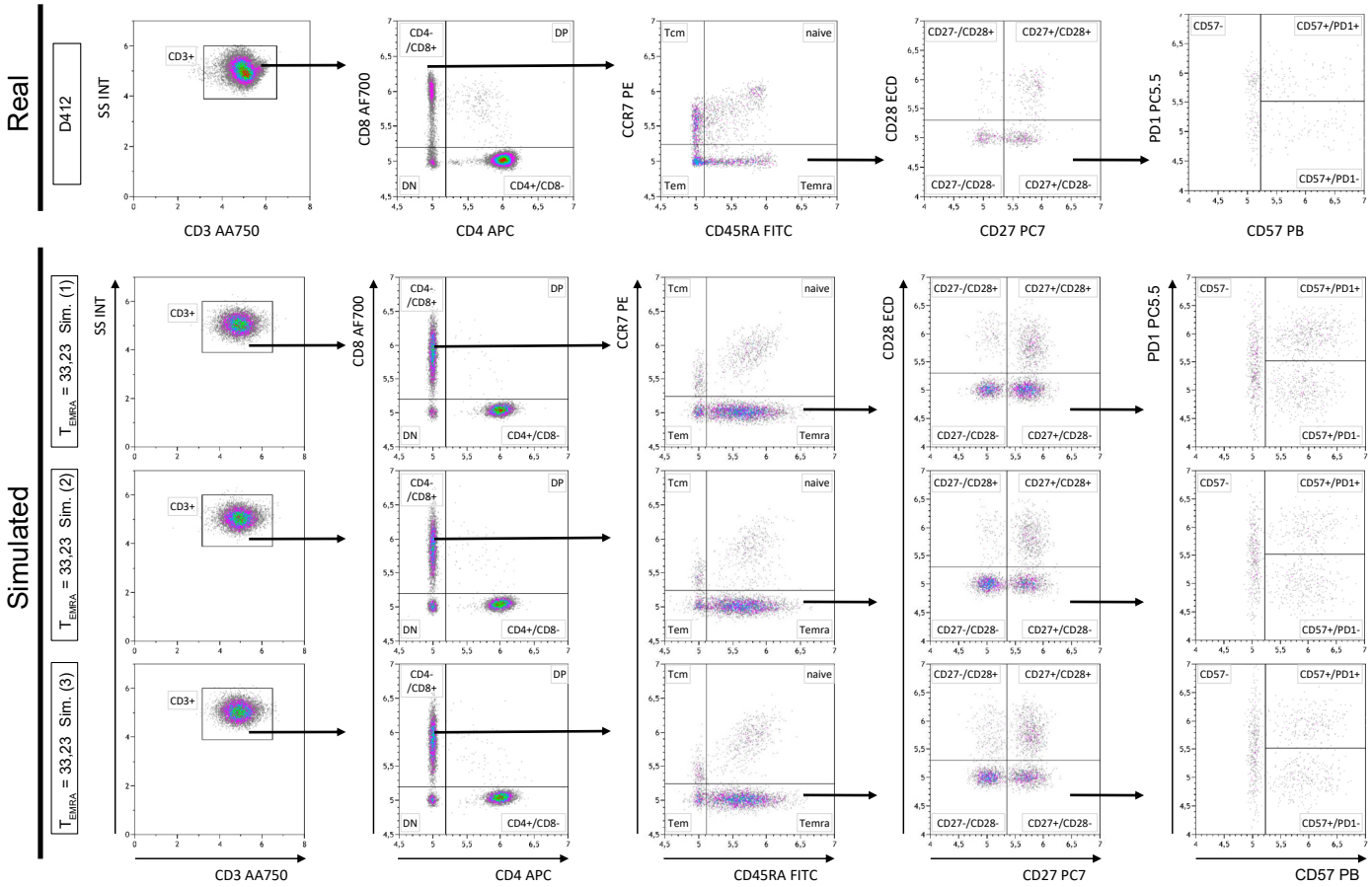
Supplementary Figure 2 *part 2/2. FlowSOM clustering results in TCR$\gamma\delta^+$ cells as a marker for hepatitis.* **(g)** TCR$\gamma\delta^+$ CD8$^+$ can be subdivided in 4 populations. Cluster 173 correspond to Population 1, characterized by low CD69 expression and low cytokine production. **(h)** Staining panel used to analyzed T cell populations in this experiment.

Supplementary Figure 3



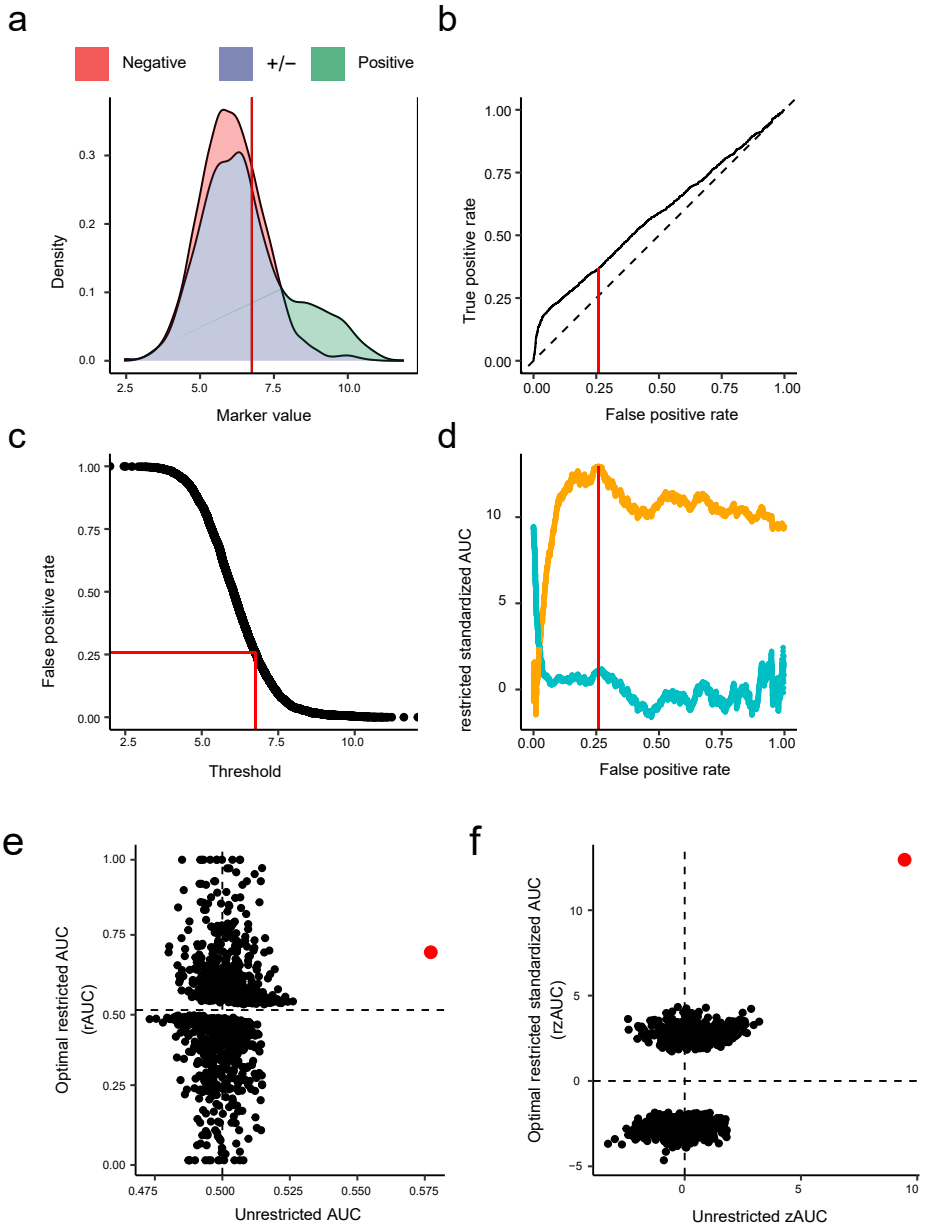
Supplementary Figure 3 *Swapped positive and negative class restriction.*
Exchanging the labels of the positive and negative classes leads to inversion of the ROC curve, but does not affect the optimal restriction. **(a)** The top-left panel shows the distribution of 2500 positive and 2500 negative samples where 20% of all positive and 2% of all negative samples are different $\mathcal{N}(9,1)$ from the majority population $\mathcal{N}(6,1)$. The optimal restriction is indicated with a red line (value = 6.8) and marker$^{\text{HIGH}}$ samples are kept. The top-right panel shows the complete ROC curve with the same restriction indicated at FPR = 0.258. The bottom-left panel relates FPR to marker values for all samples. The bottom-right panel shows the restricted standardized AUC (rzAUC) for every possible restriction calculated for marker$^{\text{HIGH}}$ and marker$^{\text{LOW}}$ samples. The optimal restriction is indicated by a red line. **(b)** This hypothetical example is identical to that shown in (a) except that the positive and negative class labels were switched when calculating the ROC curve. The top-left panel shows the distribution of 2500 negative and 2500 positive samples where 20% of all negative and 2% of all positive samples are different $\mathcal{N}(9,1)$ from the majority population $\mathcal{N}(6,1)$. The optimal restriction is indicated with a red line (value = 6.8) and marker$^{\text{HIGH}}$ samples are kept which is identical to (a). The top-right panel shows the complete ROC curve with its restriction at FPR = 0.366. The bottom-left panel relates FPR to marker values for all samples. The bottom-right panel shows the restricted standardized AUC (rzAUC) for every possible restriction calculated for marker$^{\text{HIGH}}$ and marker$^{\text{LOW}}$ samples. The optimal restriction is indicated by a red line.

Supplementary Figure 4



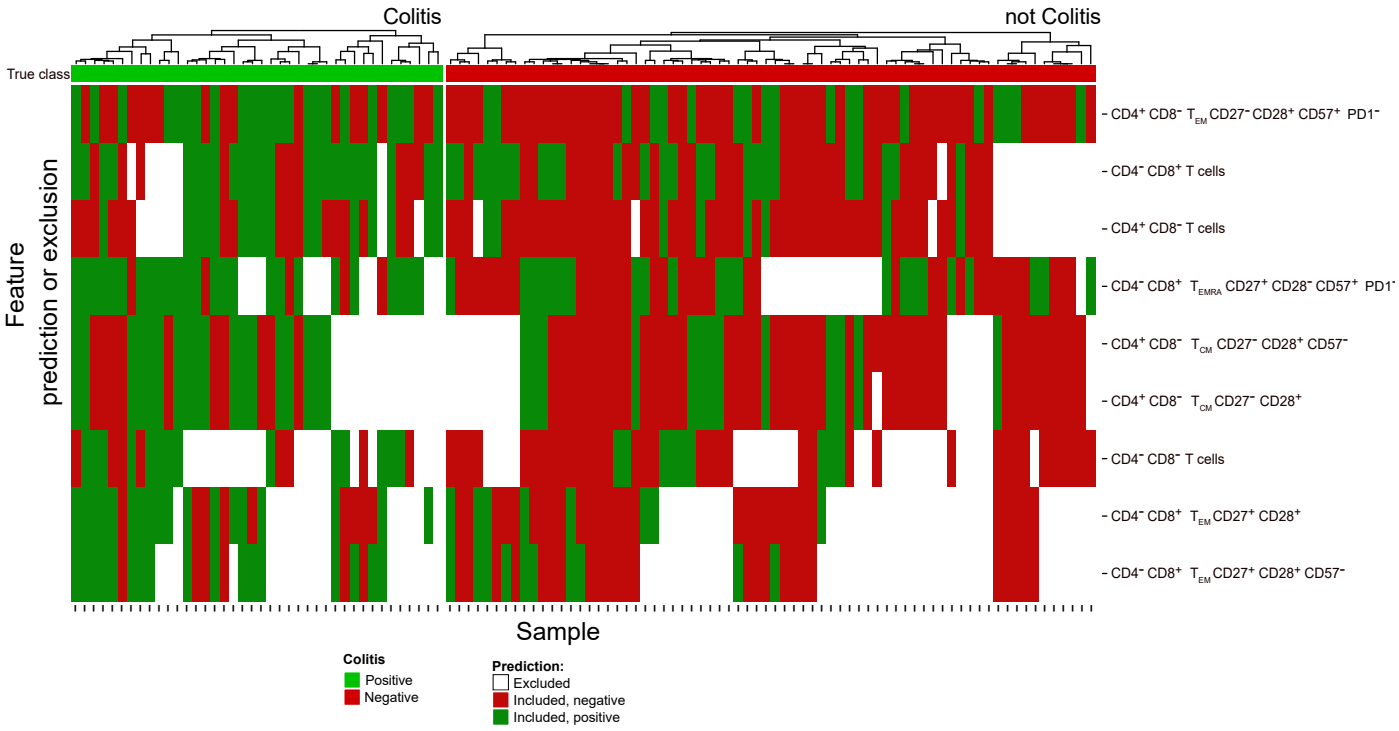
Supplementary Figure 4 *Schematic gating scheme for one real and three synthetic samples.* In this paper, our gating schema for the DURAClone IM T Cell Subsets Tube gates 1) $CD3^+$ cells, 2) splits them into quadrants according to CD4 and CD8, 3) subsequently splits $CD4^+/CD8^-$ and $CD4^-/CD8^+$ into quadrants according to CD45RA and CCR7, 4) subsequently splits every quadrant into quadrants according to CD27 and CD28, 5) subsequently splits every quadrant into $CD57^-$, $CD57^+/PD1^+$ and $CD57^+/PD1^-$. The first row shows this gating for sample D142. The next three rows show the gating for three simulated samples where $CD8^+$ T_{EMRA} were increased to have a mean of 33.23% in contrast to a baseline mean of 7.17% in 8 donors with 6 replicates each.

Supplementary Figure 5



Supplementary Figure 5 *The restricted standardized AUC is biased.*
Although the AUC of complete ROC curve is unbaised, the optimized rAUC and rzAUC are biased metrics; therefore, to compare the discriminatory performance of different markers, we introduce permutation p-values. **(a)** A simulated distribution of 2500 positive and 2500 negative samples, where 20% of all positive and 2% of all negative samples are different $\mathcal{N}(9,1)$ from the majority population $\mathcal{N}(6,1)$. The optimal restriction (value = 6.8) is indicated by a red line. **(b)** The corresponding complete ROC curve for all samples with the optimal restriction (FPR = 0.258) indicated by a red line. **(c)** A plot of marker values against FPR calculated for all samples. The optimal restriction is indicated by red lines. **(d)** The restricted standardized AUC (rzAUC) calculated for marker^{HIGH} (orange) and marker^{LOW} (blue) samples calculated for all possible restrictions. The optimal restriction is indicated by a red line. **(e)** Plot relating the AUC of the complete ROC curve to the rAUC for 10000 permutations of the positive- and negative-class labels. The red point represents the observed value for the unpermutated (i.e. correctly labelled) data. **(f)** Plot relating the standardized AUC of the complete ROC curve (zAUC) to the rzAUC for 10000 permutations of the positive- and negative-class labels. The red point represents the observed value for the unpermutated data.

Supplementary Figure 6



Supplementary Figure 6 *Significant markers according to classical and restricted analysis predicting colitis.*

The heatmap shows the significant markers for colitis based on permutation p-values for unrestricted AUC and our restriction method labelled on the right. No feature remained significant after correction for multiple testing. The markers were cell populations from the DURAClone IM T Cell Subsets Tube with our gating strategy described in Figure 4a and Supplementary Figure 4. Each further column reflects one sample, each row a feature. The samples are grouped according to their true positive (green) or negative (red) class shown in the very first row. The main matrix consists of three values: 1) Excluded according to the restriction (white/empty), 2) included and predicted positive (dark green) and 3) included and predicted negative (dark red).

Supplementary Table 1

	Training Set	Validation Set
Total number of cases	110	30
Enrollment	10/2016 - 06/2021	06/2021 - 01/202
Female	37 (33.6 %)	12 (40.0 %)
Male	73 (66.4 %)	18 (60.0 %)
Age (years)	62 (22-84)	64 (23-84)
BMI	26.6 (15.4-54.6)	28.3 (15.0-44.6)
Stage III	8 (7.3 %)	1 (3.3 %)
Stage IV	102 (92.7 %)	29 (96.7 %)
Liver metastases present	30 (27.3 %)	9 (30.0 %)
Pretreatment		
None	3 (2.7 %)	1 (3.3 %)
Surgical excision	102 (92.7%)	29 (96.7 %)
Radiosurgery	3 (2.7 %)	0 (0.0 %)
Radiation	42 (38.2 %)	6 (20.0 %)
Monotherapy	17 (15.5 %)	5 (16.7 %)
IFNa therapy	9 (8.2 %)	2 (6.7 %)
Braf/Mek inhibitor therapy	21 (19.1 %)	4 (13.3 %)
T-VEC therapy	7 (6.4 %)	0 (0.0 %)
Chemotherapy	6 (5.5 %)	3 (10.0 %)
Rounds of Ipi/Nivo		
1 round	13 (11.8 %)	5 (16.7 %)
2 rounds	24 (21.8 %)	2 (6.7 %)
3 rounds	20 (18.2 %)	9 (30.0 %)
4 rounds	53 (48.2 %)	14 (46.7 %)
Complications		
No complication	35 (31.8 %)	6 (20.0 %)
Hepatitis	48 (43.6 %)	16 (53.3 %)
Colitis	40 (36.4 %)	12 (40.0 %)
Hepatitis and Colitis	13 (11.8 %)	4 (13.3 %)

Supplementary Table 1 *Patient cohort characteristics of training and validation set.*
EDTA blood samples of 140 stage III/IV melanoma patients were collected from OCT-2016 until JAN-2023. Data from 110 training set patients have been published previously (Glehr, 2022). For age and BMI, median values were calculated. Minimum and maximum values are given in brackets. Patient cohort characteristics were obtained at start of Ipi/Nivo therapy.

Supplementary Note 2: ROC definition and probabilistic interpretation

2.1 Definition of the ROC

Let a threshold $c \in \mathbb{R}$, a continuous biomarker $Y \in \mathbb{R}$ and a grouping of samples into diseased (positive, $D = 1$) and non-diseased (negative, $D = 0$). A sample can be classified into diseased if $Y \geq c$ and into non-diseased if $Y < c$. The true positive rate (TPR) and false positive rate (FPR) at threshold c are defined as

$$\mathrm{TPR}(c) = \mathrm{P}[Y \geq c \mid D = 1], \tag{1}$$
$$\mathrm{FPR}(c) = \mathrm{P}[Y \geq c \mid D = 0]. \tag{2}$$

The ROC curve connects the TPR and FPR for all possible thresholds. We can write the set of points of the ROC curve as

$$\mathrm{ROC}_{set} = \left\{ \begin{pmatrix} \mathrm{FPR}(c) \\ \mathrm{TPR}(c) \end{pmatrix}, c \in (-\infty, \infty) \right\}. \tag{3}$$

With Equations (1) and (2), as the threshold c increases, both $\mathrm{FPR}(c)$ and $\mathrm{TPR}(c)$ decrease, see Supplementary Figure 7:

$$\lim_{c \to \infty} \mathrm{TPR}(c) = 0, \qquad \lim_{c \to -\infty} \mathrm{TPR}(c) = 1, \qquad \lim_{c \to \infty} \mathrm{FPR}(c) = 0, \qquad \lim_{c \to -\infty} \mathrm{FPR}(c) = 1, \tag{4}$$

therefore

$$\lim_{c \to \infty} \mathrm{ROC}_{set}(c) = \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix} \right\}, \qquad \lim_{c \to -\infty} \mathrm{ROC}_{set}(c) = \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\}. \tag{5}$$

The ROC curve is a monotonically increasing function. After substitution in Equation (3)

$$t := \mathrm{FPR}(c) \iff \mathrm{FPR}^{-1}(t) = c, \tag{6}$$

where c is the threshold corresponding to the false positive rate t such that $P[Y \geq c|D = 0] = t$. Here we assume that FPR is strictly monotonic such that it is invertible. Then

$$\mathrm{ROC}_{set} = \left\{ \begin{pmatrix} \mathrm{FPR}\left(\mathrm{FPR}^{-1}(t)\right) \\ \mathrm{TPR}(\mathrm{FPR}^{-1}(t)) \end{pmatrix}, t \in (0, 1) \right\}, \tag{7}$$
$$= \left\{ \begin{pmatrix} t \\ \mathrm{ROC}(t) \end{pmatrix}, t \in (0, 1) \right\}, \tag{8}$$

with the definition

$$\mathrm{ROC}(t) := \mathrm{TPR}(\mathrm{FPR}^{-1}(t)). \tag{9}$$

We simplify notation:

$$\mathrm{P}\left[Y_D \geq y\right] := \mathrm{P}\left[Y \geq y \mid D = 1\right], \tag{10}$$
$$\mathrm{P}\left[Y_{\overline{D}} \geq y\right] := \mathrm{P}\left[Y \geq y \mid D = 0\right]. \tag{11}$$

We describe the empirical survival functions from given diseased and non-diseased values

$$S_D(y) := \mathrm{P}\left[Y_D \geq y\right] = \int_y^\infty f_D(u)du \hat{=} \mathrm{TPR}(y), \tag{12}$$
$$S_{\overline{D}}(y) := \mathrm{P}\left[Y_{\overline{D}} \geq y\right] = \int_y^\infty f_{\overline{D}}(u)du \hat{=} \mathrm{FPR}(y), \tag{13}$$

where $f_D(y)$ and $f_{\overline{D}}(y)$ are the probability density functions of Y_D and $Y_{\overline{D}}$, respectively. Their derivatives are the negative probability density functions:

$$\frac{\mathrm{d}S_D(y)}{\mathrm{d}y} = -f_D(y), \qquad \frac{\mathrm{d}S_{\overline{D}}(y)}{\mathrm{d}y} = -f_{\overline{D}}(y). \tag{14}$$

In terms of the survival functions,

$$\mathrm{ROC}(t) = S_D\left(S_{\overline{D}}^{-1}(t)\right), t \in (0,1), \tag{15}$$

and following Equation (6),

$$P[Y_D \geq S_D^{-1}(t)] = t, \tag{16}$$

$$P[Y_{\overline{D}} \geq S_{\overline{D}}^{-1}(t)] = t. \tag{17}$$

We define

$$f_{D\overline{D}}(u,y) \coloneqq f_D(u) \cdot f_{\overline{D}}(y), \tag{18}$$

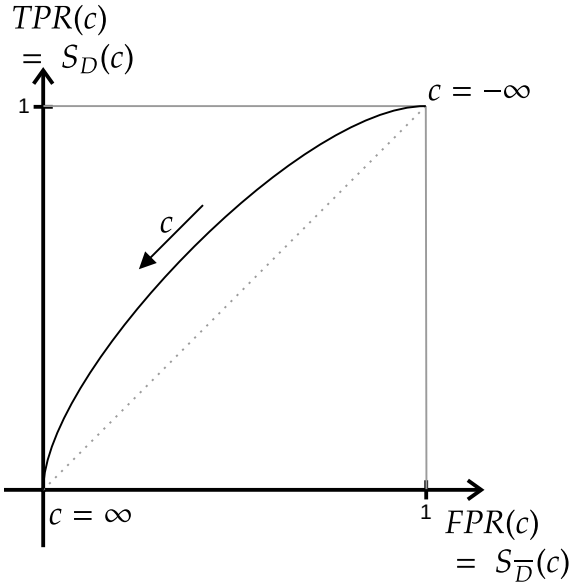
as the joint probability density function of Y_D and $Y_{\overline{D}}$ for independent random variables Y_D and $Y_{\overline{D}}$.

We further describe the following probability in terms of the empirical survival functions:

$$P[a \leq Y_X \leq b] = \int_a^b f_X(y)dy \tag{19}$$

$$= \int_a^\infty f_X(y)dy - \int_b^\infty f_X(y)dy \tag{20}$$

$$= S_X(a) - S_X(b), \qquad X \in \{D, \overline{D}\}. \tag{21}$$



Supplementary Figure 7 *ROC curve with explicit threshold notation*

We show the ROC curve using notations used during the proof. The ROC curve shows the true positive rate $TPR(c)$ or empirical survival function for diseased $S_D(c)$ on the y-axis and the false positive rate $FPR(c)$ or empirical survival function for non-diseased $S_{\overline{D}}(c)$ on the x-axis for any possible threshold c. Samples with higher values than c are classified diseased (or positive), samples with smaller or equal values are classified as non-diseased (or negative). An infinitely high positive threshold predicts all samples as negative, resulting in a true positive rate and false positive rate of 0. Conversely, an infinitely low negative threshold predicts all samples as positive, resulting in a true positive rate and false positive rate of 1.

2.2 The area under the ROC curve (AUC)

The area under the ROC curve is defined as

$$\mathrm{AUC} = \int_0^1 \mathrm{ROC}(t)\mathrm{d}t, \tag{22}$$

where a perfect ROC curve has an AUC of 1. "Perfect" corresponds to a perfectly discriminating biomarker where higher values correspond to the positive class. An uninformative biomarker has an AUC of 0.5 because $\mathrm{ROC}(t) = t$ for $t \in (0, 1)$. A perfectly discriminating biomarker with higher values corresponding to the *negative* class has an AUC of 0.

From a probabilistic point of view, the AUC equals the probability that the biomarker value of a random positive sample is higher than of a random negative sample (Bamber, 1975; Hanley and McNeil, 1982):

$$\mathrm{AUC} = \mathrm{P}\left[Y_D > Y_{\overline{D}}\right]. \tag{23}$$

Proof

$$\mathrm{AUC} = \int_0^1 \mathrm{ROC}(t)\mathrm{d}t \tag{24}$$

$$\downarrow \text{With Equation (15)}$$

$$= \int_0^1 S_D\left(S_{\overline{D}}^{-1}(t)\right)\mathrm{d}t \tag{25}$$

$$\downarrow \text{Substitute } y := S_{\overline{D}}^{-1}(t) \iff t = S_{\overline{D}}(y)$$
$$\text{With Equation (5):}$$
$$S_{\overline{D}}^{-1}(1) = -\infty$$
$$S_{\overline{D}}^{-1}(0) = \infty$$

$$= \int_{S_{\overline{D}}^{-1}(0)=\infty}^{S_{\overline{D}}^{-1}(1)=-\infty} S_D(y)S'_{\overline{D}}(y)\mathrm{d}y$$

$$\downarrow \text{With Equation (13) and Equation (14)}$$

$$= \int_{\infty}^{-\infty} S_D(y)(-f_{\overline{D}}(y))\mathrm{d}y \tag{26}$$

$$\downarrow \text{With Equation (12)}$$

$$= \int_{-\infty}^{\infty}\int_{y}^{\infty} f_D(u)\mathrm{d}u f_{\overline{D}}(y)\,\mathrm{d}y \tag{27}$$

$$= \int_{-\infty}^{\infty}\int_{y}^{\infty} f_D(u)f_{\overline{D}}(y)\mathrm{d}u\,\mathrm{d}y \tag{28}$$

$$\downarrow \text{With Equation (18)}$$

$$= \int_{-\infty}^{\infty}\int_{y}^{\infty} f_{D\overline{D}}(u,y)\mathrm{d}u\,\mathrm{d}y \tag{29}$$

$$= \mathrm{P}\left[Y_D \geq Y_{\overline{D}}\right] \tag{30}$$

Supplementary Note 3: Restriction method

3.1 Partial area under the ROC curve (pAUC)

The partial AUC (pAUC) has been defined as the AUC up to a certain false positive rate t_0 and its probabilistic correspondence has been shown (Pepe 2004, Dodd 2001):

$$\mathrm{pAUC}(t_0) = \int_0^{t_0} \mathrm{ROC}(t)\mathrm{d}t = t_0 \cdot \mathrm{P}\left[Y_D > Y_{\overline{D}} | Y_{\overline{D}} > S_{\overline{D}}^{-1}(t_0)\right], \tag{31}$$

$$\mathrm{pAUC}(t_0) = \begin{cases} \int_0^{t_0} 1\mathrm{d}t = t_0 & \text{perfect test} \\ \int_0^{t_0} t\mathrm{d}t = \frac{t_0^2}{2} & \text{uninformative test.} \end{cases} \tag{32}$$

$$\mathrm{pAUC}(t_0) = \int_0^{t_0} \mathrm{ROC}(t)\mathrm{d}t \tag{33}$$

$$\downarrow \text{With Equation (15)}$$

$$= \int_0^{t_0} S_D\left(S_{\overline{D}}^{-1}(t)\right)\mathrm{d}t \tag{34}$$

$$\downarrow \text{Substitute } y := S_{\overline{D}}^{-1}(t) \iff t = S_{\overline{D}}(y)$$
$$\text{With Equation (5): } S_{\overline{D}}^{-1}(0) = \infty$$

$$= \int_{S_{\overline{D}}^{-1}(0)=\infty}^{S_{\overline{D}}^{-1}(t_0)} S_D(y) S'_{\overline{D}}(y)\mathrm{d}y \tag{35}$$

$$\downarrow \text{With Equation (13) and Equation (14)}$$

$$= \int_{\infty}^{S_{\overline{D}}^{-1}(t_0)} S_D(y)(-f_{\overline{D}}(y))\mathrm{d}y$$

$$\downarrow \text{With Equation (12)}$$

$$= \int_{S_{\overline{D}}^{-1}(t_0)}^{\infty} \int_y^{\infty} f_D(u)\mathrm{d}u f_{\overline{D}}(y)\,\mathrm{d}y \tag{36}$$

$$= \int_{S_{\overline{D}}^{-1}(t_0)}^{\infty} \int_y^{\infty} f_D(u) f_{\overline{D}}(y)\mathrm{d}u\,\mathrm{d}y \tag{37}$$

$$\downarrow \text{With Equation (18)}$$

$$= \int_{S_{\overline{D}}^{-1}(t_0)}^{\infty} \int_y^{\infty} f_{D\overline{D}}(u, y)\mathrm{d}u\,\mathrm{d}y \tag{38}$$

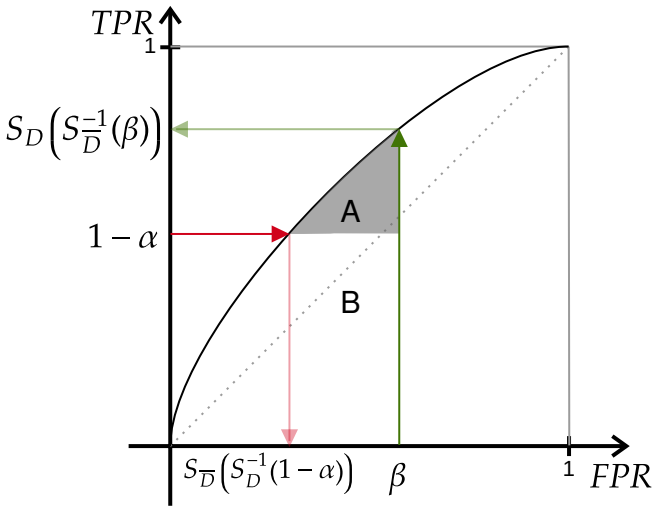
$$= \mathrm{P}\left[Y_D \geq Y_{\overline{D}}, Y_{\overline{D}} \geq S_{\overline{D}}^{-1}(t_0)\right] \tag{39}$$

$$= \mathrm{P}\left[Y_{\overline{D}} \geq S_{\overline{D}}^{-1}(t_0)\right] \mathrm{P}\left[Y_D \geq Y_{\overline{D}} | Y_{\overline{D}} \geq S_{\overline{D}}^{-1}(t_0)\right] \tag{40}$$

$$\downarrow \text{with Equation (16): } \mathrm{P}\left[Y_{\overline{D}} \geq S_{\overline{D}}^{-1}(t_0)\right] = t_0$$

$$\mathrm{pAUC}(t_0) = t_0 \cdot \mathrm{P}\left[Y_D \geq Y_{\overline{D}} | Y_{\overline{D}} \geq S_{\overline{D}}^{-1}(t_0)\right] \tag{41}$$

3.2 Two-way partial area under the ROC curve



Supplementary Figure 8 *Two-way partial area under the ROC curve.*
The two-way partial area under the ROC curve is defined as the area under the ROC curve between a maximum false positive rate β and a minimum true positive rate $1-\alpha$. It can be calculated by calculating the usual area under the curve between the two corresponding false positive rates $\left[S_{\overline{D}}(S_D^{-1}(1-\alpha)), \beta\right]$ and subtracting the area of the rectangle B.

The concept of partial area under the curve (pAUC) has been expanded in recent years to include two-way partial AUCs (Yang 2017 and Yang 2021), which allow for the calculation of the area under the curve between an upper limit for the false positive rate and a lower limit for the true positive rate ($S_{\overline{D}}(t) \leq \beta$ and $S_D(t) \geq 1-\alpha$). This area, shown in Supplementary Figure 8 as shaded area A, can be written as

$$\mathrm{AUC}_\alpha^\beta = (Area_A + Area_B) - Area_B \tag{42}$$

$$= \int_{S_{\overline{D}}\left(S_D^{-1}(1-\alpha)\right)}^{\beta} S_D(S_{\overline{D}}^{-1}(t))\mathrm{d}t - (1-\alpha)\left(\beta - S_{\overline{D}}\left(S_D^{-1}(1-\alpha)\right)\right) \tag{43}$$

↓ See following proof.

$$= P\left[Y_D > Y_{\overline{D}}, Y_D \leq S_D^{-1}(1-\alpha), Y_{\overline{D}} \geq S_{\overline{D}}^{-1}(\beta)\right] \tag{44}$$

$$= P\left[S_D^{-1}(1-\alpha) \geq Y_D > Y_{\overline{D}} \geq S_{\overline{D}}^{-1}(\beta)\right]. \tag{45}$$

Our method uses two special cases of $\mathrm{AUC}_\alpha^\beta$: The bottom-left ($AUC_{high}$, green) and top-right ($\mathrm{AUC}_{low}$, red) part of the AUC, shown in Supplementary Figure 9.

Proof

With

$$y_0 = S_{\overline{D}}^{-1}(t_0), \qquad y_1 = S_{\overline{D}}^{-1}(t_1), \tag{46}$$

we have

$$\int_{t_1}^{t_0} \mathrm{ROC}(t)dt = \int_{t_1}^{t_0} S_D(\underbrace{S_{\overline{D}}^{-1}(t)}_{=y})dt = \int_{y_1}^{y_0} S_D(y)S'_{\overline{D}}(y)dy = \int_{y_0}^{y_1} S_D(y)f_{\overline{D}}(y)dy\,. \tag{47}$$

From

$$S_D(y) = \int_y^{\infty} f_D(u)du = \int_y^{y_1} f_D(u)du + \int_{y_1}^{\infty} f_D(u)du = \int_y^{y_1} f_D(u)du + S_D(y_1) \tag{48}$$

we obtain

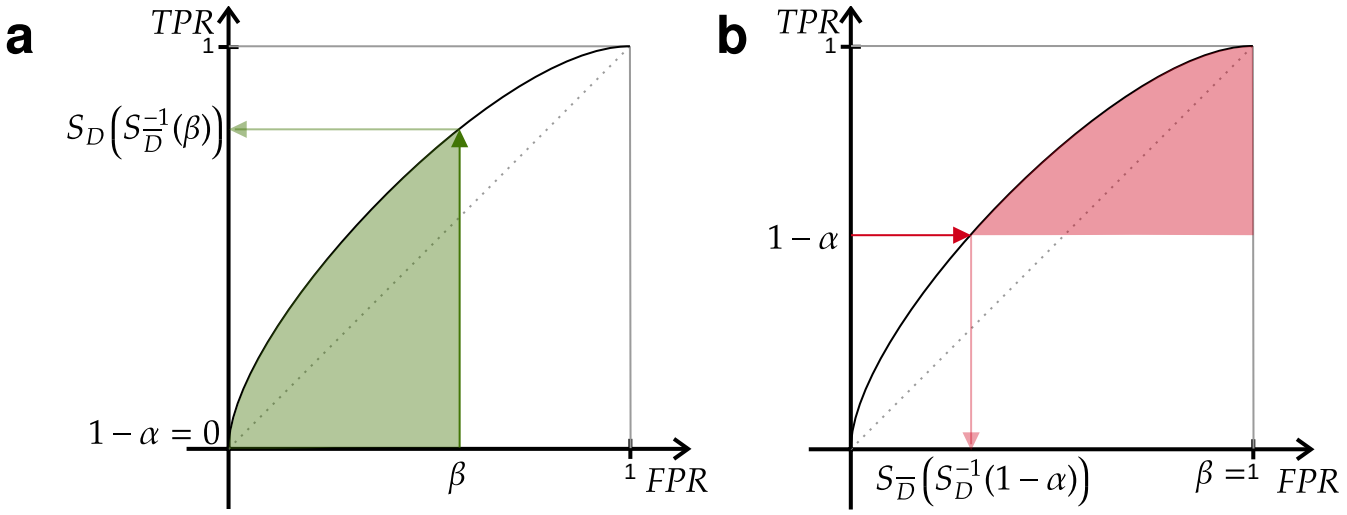
$$\begin{aligned}
\int_{t_1}^{t_0} \mathrm{ROC}(t)dt &= \int_{y_0}^{y_1} S_D(y)f_{\overline{D}}(y)dy && (49)\\
&= \int_{y_0}^{y_1} \left[\int_y^{y_1} f_D(u)du + S_D(y_1)\right] f_{\overline{D}}(y)dy && (50)\\
&= \int_{y_0}^{y_1}\int_y^{y_1} f_D(u)du f_{\overline{D}}(y)dy + S_D(y_1)\int_{y_0}^{y_1} f_{\overline{D}}(y)dy\\
&= \int_{y_0}^{y_1}\int_y^{y_1} f_D(u)du f_{\overline{D}}(y)dy + S_D(y_1)\left(S_{\overline{D}}(y_0) - S_{\overline{D}}(y_1)\right)\\
&= P[y_1 \geq Y_D > Y_{\overline{D}} \geq y_0] + S_D(y_1)\,(t_0 - t_1)\,. && (51)
\end{aligned}$$

For our application, use

$$t_0 = \beta\,, \quad y_0 = S_{\overline{D}}^{-1}(\beta)\,, \quad t_1 = S_{\overline{D}}(S_D^{-1}(1-\alpha))\,, \quad y_1 = S_D^{-1}(1-\alpha)\,, \quad S_D(y_1) = (1-\alpha)\,, \tag{52}$$

to obtain the two-way partial AUC

$$\mathrm{AUC}_{\alpha}^{\beta} = \int_{S_{\overline{D}}\left(S_D^{-1}(1-\alpha)\right)}^{\beta} \mathrm{ROC}(t)\mathrm{d}t - (1-\alpha)\left(\beta - S_{\overline{D}}\left(S_D^{-1}(1-\alpha)\right)\right). \tag{53}$$



Supplementary Figure 9 *ROC curves showing two-way partial AUCs corresponding to* $marker^{HIGH}$ *and* $marker^{LOW}$ *samples used in our restriction method.*
Partial AUCs using the nomenclature from Supplementary Figure 8. **(a)** Two-way partial AUC with $\alpha = 1$ and a maximum false positive rate β. β corresponds to a specific threshold through $c = S_{\overline{D}}^{-1}(\beta)$. Therefore this green area corresponds to the AUC calculated for the complete ROC curve including $marker^{HIGH}$ samples. **(b)** Two-way partial AUC with $\beta = 1$ and a minimum true positive rate of $1-\alpha$. $1-\alpha$ corresponds to a specific threshold through $c = S_D^{-1}(1-\alpha)$. Therefore this red area corresponds to the AUC calculated for the complete ROC curve including $marker^{LOW}$ samples.

3.2.1 AUC_{high}

The left part of the area under the curve up to a false positive rate β is identical to the pAUC described earlier (Equation (31))

$$\mathrm{AUC}_{high}(\beta) = \mathrm{AUC}^{\beta}_{\underset{\equiv 1}{\alpha}} = \tag{54}$$

$$\downarrow \text{Equation (43)}$$

$$= \int_{S_{\overline{D}}\left(S_D^{-1}(1-\underset{\equiv 1}{\alpha})\right)}^{\beta} \mathrm{ROC}(t)\mathrm{d}t - (1-\underset{\equiv 1}{\alpha})\left(\beta - S_{\overline{D}}\left(S_D^{-1}(1-\underset{\equiv 1}{\alpha})\right)\right) \tag{55}$$

$$= \int_{S_{\overline{D}}\left(S_D^{-1}(0)\right)}^{\beta} \mathrm{ROC}(t)\mathrm{d}t \tag{56}$$

$$\downarrow \text{Equation (5)}$$

$$= \int_{S_{\overline{D}}(\infty)}^{\beta} \mathrm{ROC}(t)\mathrm{d}t \tag{57}$$

$$= \int_0^{\beta} \mathrm{ROC}(t)\mathrm{d}t \overset{Equation\ (31)}{=} pAUC(\beta). \tag{58}$$

$$\mathrm{AUC}_{high}(\beta) =$$

$$\downarrow \text{Equation (44)}$$

$$= P\left[Y_D > Y_{\overline{D}}, Y_D \leq S_D^{-1}(1 - \underset{\equiv 1}{\alpha}), Y_{\overline{D}} \geq S_{\overline{D}}^{-1}(\beta)\right] \tag{59}$$

$$= P\left[Y_D > Y_{\overline{D}}, Y_D \leq S_D^{-1}(0), Y_{\overline{D}} \geq S_{\overline{D}}^{-1}(\beta)\right] \tag{60}$$

$$= P\left[Y_D > Y_{\overline{D}}, Y_D \leq \infty, Y_{\overline{D}} \geq S_{\overline{D}}^{-1}(\beta)\right] \tag{61}$$

$$= P\left[Y_D > Y_{\overline{D}}, Y_{\overline{D}} \geq S_{\overline{D}}^{-1}(\beta)\right] \tag{62}$$

$$\downarrow \text{Equation (39) and } t_0 := \beta$$

$$= pAUC(\beta). \tag{63}$$

3.2.2 AUC_{low}

The right part of the area under the curve with at least a true positive rate of $1 - \alpha$

$$\mathrm{AUC}_{low}(\alpha) = \mathrm{AUC}_{\alpha}^{\beta \equiv 1} = \tag{64}$$

$$\downarrow \text{Equation (43)}$$

$$= \int_{S_{\overline{D}}\left(S_D^{-1}(1-\alpha)\right)}^{\underset{\equiv 1}{\beta}} \mathrm{ROC}(t)\mathrm{d}t - (1-\alpha)\left(\underset{\equiv 1}{\beta} - S_{\overline{D}}\left(S_D^{-1}(1-\alpha)\right)\right) \tag{65}$$

$$= \int_{S_{\overline{D}}\left(S_D^{-1}(1-\alpha)\right)}^{1} \mathrm{ROC}(t)\mathrm{d}t - (1-\alpha)\left(1 - S_{\overline{D}}\left(S_D^{-1}(1-\alpha)\right)\right). \tag{66}$$

$$\mathrm{AUC}_{low}(\alpha) =$$

$$\downarrow \text{Equation (44)}$$

$$= P\left[Y_D > Y_{\overline{D}}, Y_D \leq S_D^{-1}(1-\alpha), Y_{\overline{D}} \geq S_{\overline{D}}^{-1}(\underset{\equiv 1}{\beta})\right] \tag{67}$$

$$\downarrow \text{Equation (5)}$$

$$= P\left[Y_D > Y_{\overline{D}}, Y_D \leq S_D^{-1}(1-\alpha), Y_{\overline{D}} \geq -\infty\right] \tag{68}$$

$$= P\left[Y_D > Y_{\overline{D}}, Y_D \leq S_D^{-1}(1-\alpha)\right] \tag{69}$$

$$= P\left[Y_D \leq S_D^{-1}(1-\alpha)\right] P\left[Y_D > Y_{\overline{D}} | Y_D \leq S_D^{-1}(1-\alpha)\right] \tag{70}$$

$$= (1-\alpha) \cdot P\left[Y_D > Y_{\overline{D}} | Y_D \leq S_D^{-1}(1-\alpha)\right]. \tag{71}$$

3.3 Restricted sample space

3.3.1 ROC for restricted sample space

We restrict our sample space to

$$y_0 \leq Y_X \leq y_1 \qquad \text{for both} \quad X \in \{D, \overline{D}\}\,. \tag{72}$$

This restriction leads to modified probability density functions

$$\tilde{f}_X(y) = \begin{cases} \frac{1}{Z_X} f_X(y) & \text{for } y \in [y_0, y_1]\,, \\ 0 & \text{otherwise}\,. \end{cases} \tag{73}$$

Throughout, explicit dependence on y_0, y_1 is omitted to increase readability but always implied. The (y-independent) constants

$$Z_X = \int_{y_0}^{y_1} f_X(y)dy = S_X(y_0) - S_X(y_1) = P[y_1 \geq Y_X \geq y_0] \tag{74}$$

ensure proper normalization of the densities,

$$\int_{-\infty}^{\infty} \tilde{f}_X(y)dy = \int_{y_0}^{y_1} \tilde{f}_X(y)dy = \frac{1}{Z_X}\int_{y_0}^{y_1} f_X(y)dy = 1\,. \tag{75}$$

This means that $\tilde{f}_X$ can be interpreted as a conditional probability density function for Y_X, conditioned on the restriction $y_0 \leq Y_X \leq y_1$. For quantities defined on the restricted space we use $\widetilde{(\cdot)}$.

On the restricted space, true and false positive rates can be determined from the corresponding densities $\tilde{f}_D$ and $\tilde{f}_{\overline{D}}$ in the usual way,

$$\widetilde{S}_X(y) = \int_y^{\infty} \tilde{f}_X(u)du = \begin{cases} 1 & \text{for } y < y_0\,, \\ \int_y^{y_1} \tilde{f}_X(u)du & \text{for } y_0 \leq y \leq y_1\,, \\ 0 & \text{for } y_1 < y\,. \end{cases} \tag{76}$$

For the non-trivial case $y_0 \leq y \leq y_1$, we therefore obtain

$$\begin{aligned} \widetilde{S}_X(y) &= \int_y^{y_1} \tilde{f}_X(u)du = \frac{1}{Z_X}\int_y^{y_1} f_X(u)du = \frac{S_X(y) - S_X(y_1)}{S_X(y_0) - S_X(y_1)} = \frac{P[y \leq Y_X \leq y_1]}{P[y_0 \leq Y_X \leq y_1]} \\ &= \frac{P[y_0 \leq y \leq Y_X \leq y_1]}{P[y_0 \leq Y_X \leq y_1]} = \frac{P[y \leq Y_X\,,\, y_0 \leq Y_X \leq y_1]}{P[y_0 \leq Y_X \leq y_1]} = P[y \leq Y_X \,|\, y_0 \leq Y_X \leq y_1]\,, \end{aligned} \tag{77}$$

where the probabilities $P[.]$ are defined w.r.t the original (unrestricted) density functions f_X. Equation (77) can be rewritten as

$$\widetilde{S}_X(y) = \frac{1}{S_X(y_0) - S_X(y_1)} S_X(y) - \frac{S_X(y_1)}{S_X(y_0) - S_X(y_1)} = \gamma_X S_X(y) - \delta_X \tag{78}$$

with (y-independent) constants

$$\gamma_X = \frac{1}{Z_X} = \frac{1}{S_X(y_0) - S_X(y_1)} \qquad \text{and} \qquad \delta_X = \frac{S_X(y_1)}{S_X(y_0) - S_X(y_1)} = \gamma_X S_X(y_1)\,. \tag{79}$$

This establishes a one-to-one map between the unrestricted rates $S_X(y) \in [S_X(y_1), S_X(y_0)]$ and their restricted counterparts $\widetilde{S}_X(y) \in [0, 1]$, through a simple shift (by δ_X) and rescaling (with γ_X).

From the rate functions $\widetilde{S}_X$, we obtain the corresponding ROC function on the restricted space,

$$\mathrm{rROC}(t) = \widetilde{S}_D\left(\widetilde{S}_{\overline{D}}^{-1}(t)\right), \quad t \in [0,1]\,. \tag{80}$$

Note that the interval $[0,1]$ for t corresponds to thresholds $y \in [y_0, y_1]$. As the identity Equation (78) applies to both $X \in \{D, \overline{D}\}$, we have a one-to-one map between points on the original ROC curve obtained from thresholds $y \in [y_0, y_1]$ and points on the rROC curve for the restricted densities:

$$\begin{pmatrix} S_{\overline{D}}(y) \\ S_D(y) \end{pmatrix} \quad \mapsto \quad \begin{pmatrix} \widetilde{S}_{\overline{D}}(y) \\ \widetilde{S}_D(y) \end{pmatrix} = \begin{pmatrix} \gamma_{\overline{D}} S_{\overline{D}}(y) \\ \gamma_D S_D(y) \end{pmatrix} - \begin{pmatrix} \delta_{\overline{D}} \\ \delta_D \end{pmatrix}. \tag{81}$$

In particular

$$\begin{pmatrix} S_{\overline{D}}(y_0) \\ S_D(y_0) \end{pmatrix} \quad \mapsto \quad \begin{pmatrix} \widetilde{S}_{\overline{D}}(y_0) \\ \widetilde{S}_D(y_0) \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \tag{82}$$

$$\begin{pmatrix} S_{\overline{D}}(y_1) \\ S_D(y_1) \end{pmatrix} \quad \mapsto \quad \begin{pmatrix} \widetilde{S}_{\overline{D}}(y_1) \\ \widetilde{S}_D(y_1) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{83}$$

This means that the rROC curve for the restricted sample space can be obtained from the original ROC curve by simply cutting out a rectangle corresponding to the intervals $[S_{\overline{D}}(y_1), S_{\overline{D}}(y_0)]$ and $[S_D(y_1), S_D(y_0)]$ and rescaling both axes to $[0,1]$ afterwards.

3.3.2 AUC for restricted sample space

Since we have (see Equation (77))

$$\widetilde{S}_D(y) = P[y \le Y_D \,|\, y_0 \le Y_D \le y_1]\,, \tag{84}$$

$$\widetilde{S}_{\overline{D}}(y) = P[y \le Y_{\overline{D}} \,|\, y_0 \le Y_{\overline{D}} \le y_1]\,, \tag{85}$$

the area under the rROC curve is given by

$$\mathrm{rAUC} = P[Y_{\overline{D}} \le Y_D \,|\, y_0 \le Y \le y_1]\,. \tag{86}$$

By definition, we have

$$\begin{aligned} P[y_0 \le Y \le y_1] &= P[y_0 \le Y_D \le y_1] P[y_0 \le Y_{\overline{D}} \le y_1] = Z_D Z_{\overline{D}} \\ &= \left(S_D(y_0) - S_D(y_1)\right)\left(S_{\overline{D}}(y_0) - S_{\overline{D}}(y_1)\right). \end{aligned} \tag{87}$$

For rAUC we then obtain

$$\mathrm{rAUC} = \frac{P[Y_{\overline{D}} \le Y_D\,,\, y_0 \le Y \le y_1]}{P[y_0 \le Y \le y_1]} \tag{88}$$

$$= \frac{P[y_0 \le Y_{\overline{D}} \le Y_D \le y_1]}{Z_D Z_{\overline{D}}} \tag{89}$$

$$= \frac{P[y_0 \le Y_{\overline{D}} \le Y_D \le y_1]}{\left(S_D(y_0) - S_D(y_1)\right)\left(S_{\overline{D}}(y_0) - S_{\overline{D}}(y_1)\right)}. \tag{90}$$

3.3.3 Restricted AUC in terms of true and false positive rates

We can write Equation (86) in terms of a minimum true positive rate $1-\alpha$ and a maximum false positive rate β by defining

$$y_0 := S_{\overline{D}}^{-1}(\beta)\,, \qquad y_1 := S_D^{-1}(1-\alpha), \tag{91}$$

to obtain

$$\mathrm{rAUC}_\alpha^\beta := P[Y_D > Y_{\overline{D}} \,|\, S_{\overline{D}}^{-1}(\beta) \le Y \le S_D^{-1}(1-\alpha)], \tag{92}$$

where $S_{\overline{D}}^{-1}(\beta) \le Y \le S_D^{-1}(1-\alpha)$ is the condition that all values (Y_D and $Y_{\overline{D}}$) are between $S_{\overline{D}}^{-1}(\beta)$ and $S_D^{-1}(1-\alpha)$.
With Equation (87) we get a scaling factor

$$P[y_0 \le Y \le y_1] = \Big(S_D(S_{\overline{D}}^{-1}(\beta)) - (1-\alpha)\Big)\Big(\beta - S_{\overline{D}}(S_D^{-1}(1-\alpha))\Big) = \mathrm{scaling}_\alpha^\beta\,. \tag{93}$$

This means that with Equation (45) and Equation (90), the restricted AUC can *also* be obtained from the original two-way partial AUC by dividing by this scaling factor:

$$\mathrm{rAUC}_\alpha^\beta = \frac{\mathrm{AUC}_\alpha^\beta}{\mathrm{scaling}_\alpha^\beta}\,. \tag{94}$$

This is directly visible from Supplementary Figure 8 where a rectangle spanning across area A has a width of

$$w = \beta - S_{\overline{D}}(S_D^{-1}(1-\alpha)), \tag{95}$$

and a height of

$$h = S_D(S_{\overline{D}}^{-1}(\beta)) - (1-\alpha) = \mathrm{ROC}(\beta) - (1-\alpha), \tag{96}$$

which results in the scaling factor $\mathrm{scaling}_\alpha^\beta$.

3.4 Restriction

Until now, $\mathrm{rAUC}_{\alpha}^{\beta}$ is defined in terms of maximum false positive rate β and minimum true positive rate $1-\alpha$. Alternatively, we here introduce a "restriction" $r \in \mathbb{R}$ which splits the data into high and low parts where $\alpha := 1 - S_D(r)$ and $\beta := S_{\overline{D}}(r)$, respectively.

3.4.1 $\mathrm{rAUC}_{high}(r)$

$$\mathrm{rAUC}_{high}(r) := \mathrm{rAUC}_{\alpha\equiv 1}^{\beta\equiv S_{\overline{D}}(r)} = \tag{97}$$

$$\downarrow \text{Equation (94)}$$

$$= \mathrm{AUC}_{high}(S_{\overline{D}}(r)) \cdot \frac{1}{S_{\overline{D}}(r) - S_{\overline{D}}\left(S_D^{-1}(1-1)\right)} \cdot \frac{1}{S_D\left(S_{\overline{D}}^{-1}\left(S_{\overline{D}}(r)\right)\right) - (1-1)} \tag{98}$$

$$\downarrow \text{Equation (5): } S_D^{-1}(0) = \infty \text{ and } S_{\overline{D}}(\infty) = 0$$

$$= \mathrm{AUC}_{high}(S_{\overline{D}}(r)) \cdot \frac{1}{S_{\overline{D}}(r)} \cdot \frac{1}{S_D(r)}. \tag{99}$$

3.4.2 $\mathrm{rAUC}_{low}(r)$

$$\mathrm{rAUC}_{low}(r) := \mathrm{rAUC}_{\alpha\equiv 1-S_D(r)}^{\beta\equiv 1} = \tag{100}$$

$$\downarrow \text{Equation (94)}$$

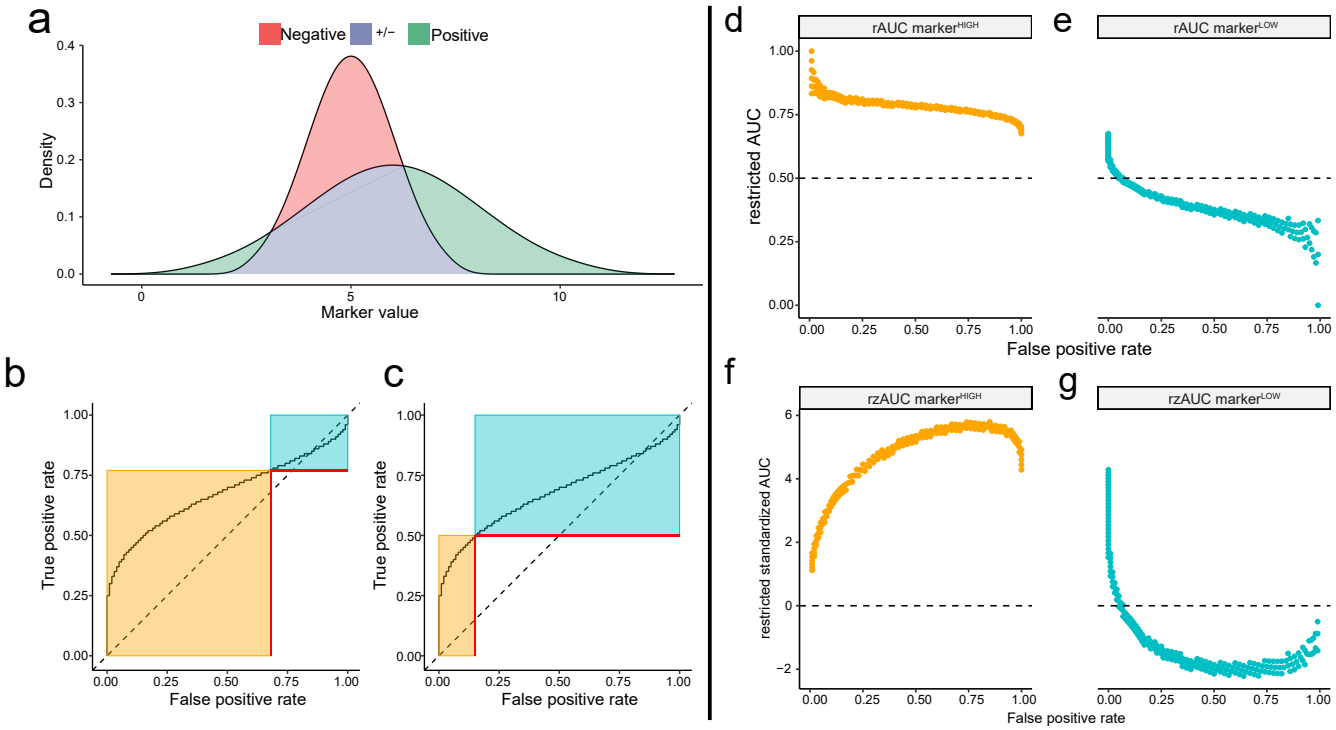
$$= \mathrm{AUC}_{low}(1-S_D(r)) \cdot \frac{1}{1 - S_{\overline{D}}\left(S_D^{-1}(1-(1-S_D(r)))\right)} \cdot \frac{1}{S_D\left(S_{\overline{D}}^{-1}(1)\right) - (1-(1-S_D(r)))} \tag{101}$$

$$\downarrow \text{Equation (5): } S_{\overline{D}}^{-1}(1) = -\infty \text{ and } S_D(-\infty) = 1$$

$$= \mathrm{AUC}_{low}(1-S_D(r)) \cdot \frac{1}{1 - S_{\overline{D}}\left(S_D^{-1}(S_D(r))\right)} \cdot \frac{1}{1-(1-(1-S_D(r)))} \tag{102}$$

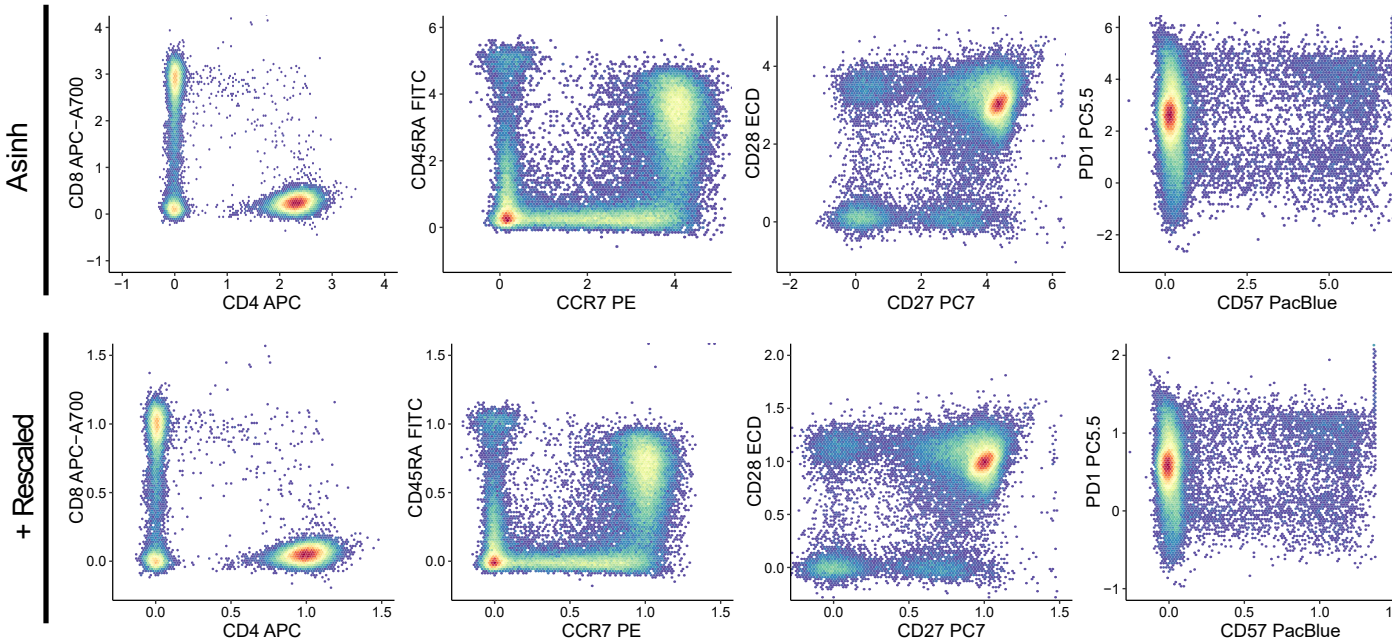
$$= \mathrm{AUC}_{low}(1-S_D(r)) \cdot \frac{1}{1 - S_{\overline{D}}(r)} \cdot \frac{1}{1-S_D(r)}. \tag{103}$$

Supplementary Figure 10



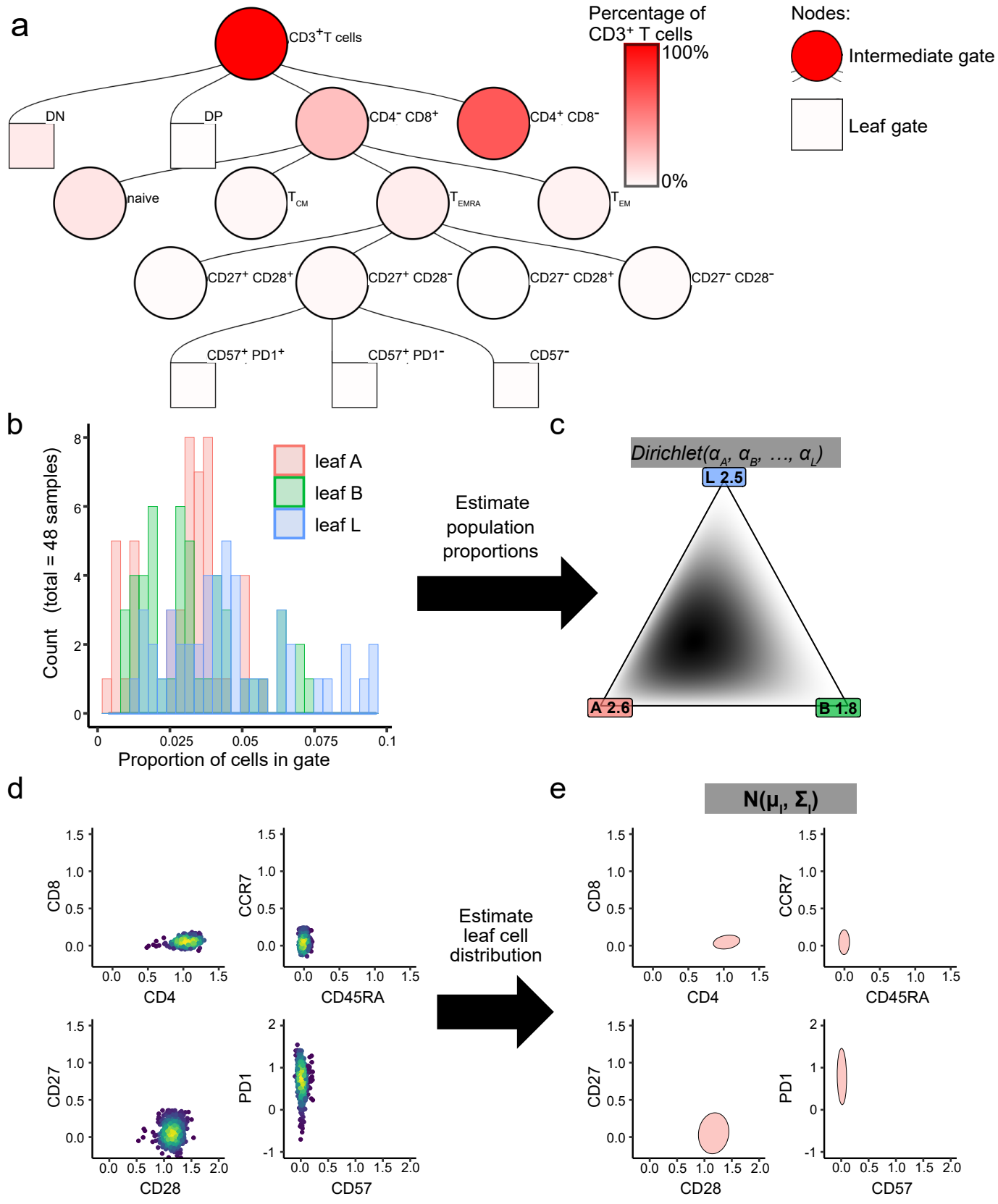
Supplementary Figure 10 *Densities, ROC curves, restricted AUCs and restricted standardized AUCs.* **(a)** A simulated distribution of 100 negative $\mathcal{N}(5,1)$ and 100 positive $\mathcal{N}(6,2)$ samples. **(b)** The corresponding right-skewed complete ROC curve showing two possible restriction levels, which illustrates that a restriction can be made at every sample value. Each restriction divides the samples into sets of marker$^{\text{HIGH}}$ (orange) and marker$^{\text{LOW}}$ (blue) samples. **(c)** See (b). **(d)** Restricted AUC (rAUC) of marker$^{\text{HIGH}}$ samples against the FPR calculated for every possible restriction. **(e)** Restricted AUC (rAUC) of marker$^{\text{LOW}}$ samples against the FPR calculated for every possible restriction. **(f)** Plots of the restricted standardized AUC (rzAUC) against the FPR calculated for every possible restriction on marker$^{\text{HIGH}}$ samples. **(g)** Plots of the restricted standardized AUC (rzAUC) against the FPR calculated for every possible restriction on marker$^{\text{LOW}}$ samples.

Supplementary Figure 11



Supplementary Figure 11 *Pre-processing of flow cytometry data.*
Following a conventional workflow, an experienced, blinded operator performed sample-wise manual recompensation of flow cytometry data using Kaluza software. Cell feature-wise asinh-transformations and rescaling were then applied. Here, we illustrate our data pre-processing with a single sample. The upper row shows all events after cell feature-wise asinh transformation. The lower row shows the same sample after rescaling.

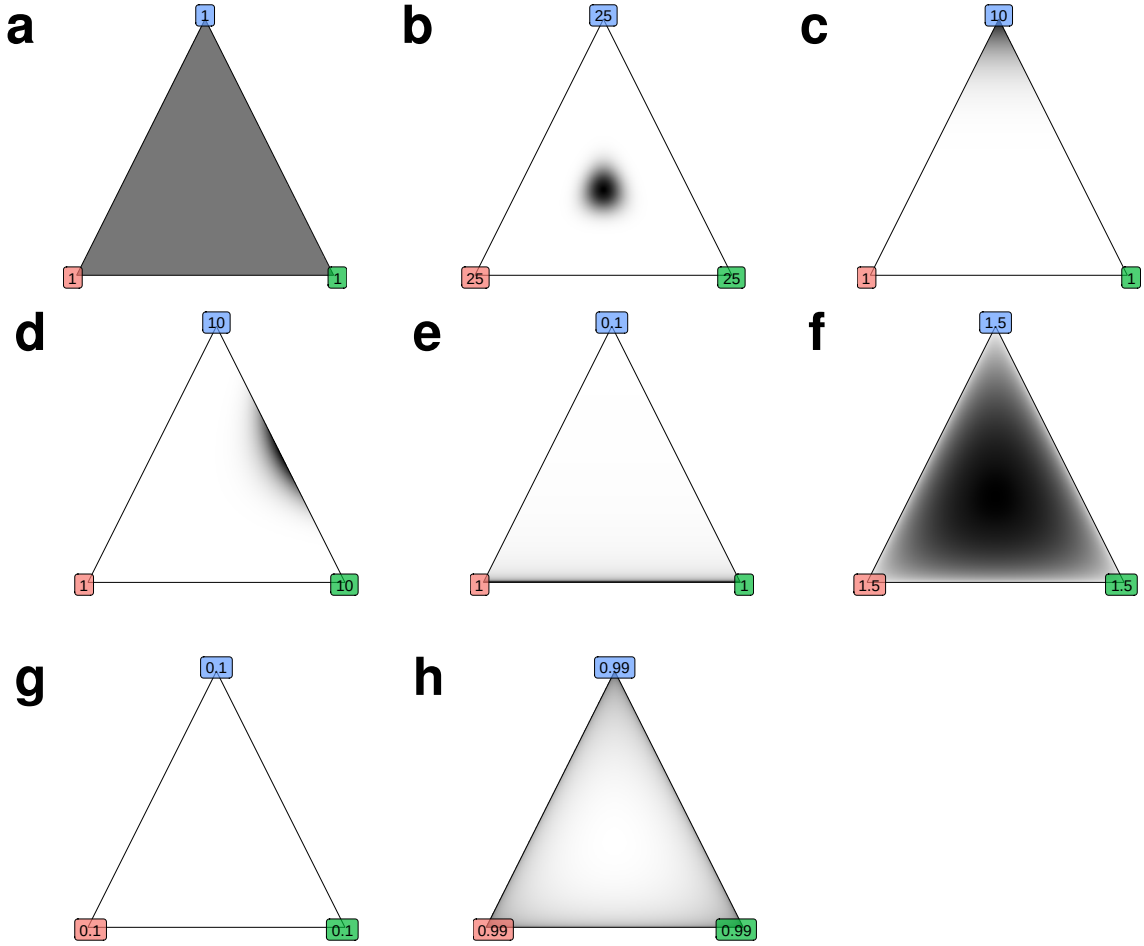
Supplementary Figure 12



Supplementary Figure 12 *Estimating cell population parameters as basis for synthesizing realistic flow cytometry data.*

Supplementary Figure 12 *Estimating cell population parameters as basis for synthesizing realistic flow cytometry data.* **(a)** A hierarchical gating tree used to classify peripheral blood T cell subsets. Within this hierarchy, intermediate gates (circles) contain the cells that are categorized into end-nodes by leaf gates (rectangles). Typical frequencies of cells in all gates were estimated from 48 samples, which represented 6 replicate stainings from 8 healthy donors. Gates are coloured according to the percentage of all cells they encompass. **(b)** A histogram showing the number of cells from all samples in three arbitrarily selected leaf gates. Leaf A (red) is the $CD27^-$ $CD28^+$ $CD57^-$ $CD4^+$ T_{EM} cell gate. Leaf B (green) is the $CD27^+$ $CD28^-$ $CD57^-$ $CD8^+$ T_{EMRA} cell gate. Leaf L (blue) is the $CD27^-$ $CD28^+$ $CD57^-$ $CD4^+$ T_{CM} cell gate. **(c)** A plot showing the corresponding density of the Dirichlet distribution for example leaf gates A, B and L. Black shading indicates higher probabilities. **(d)** 2-dimensional scatter plots showing the distribution of cell-surface protein expression (i.e. features) used to define cells that fall within leaf A. Higher cell densities are indicated with yellow shading. **(e)** Expression values for all 8 features were used to estimate a multivariate normal distribution for every leaf. The resulting leaf cell distribution for leaf A is depicted as shaded areas representing normal distributions ± 1 standard deviation.

Supplementary Figure 13



Supplementary Figure 13 *Dirichlet probablity density examples for K=3.*
Each subfigure shows the probability density for different parameters α_1, α_2 and α_3 from a Dirichlet distribution $p(\mathbf{p}) \sim Dir(\alpha_1, \alpha_2, \alpha_3)$. Parameter values are shown at the corners and also in Supplementary Table 2.

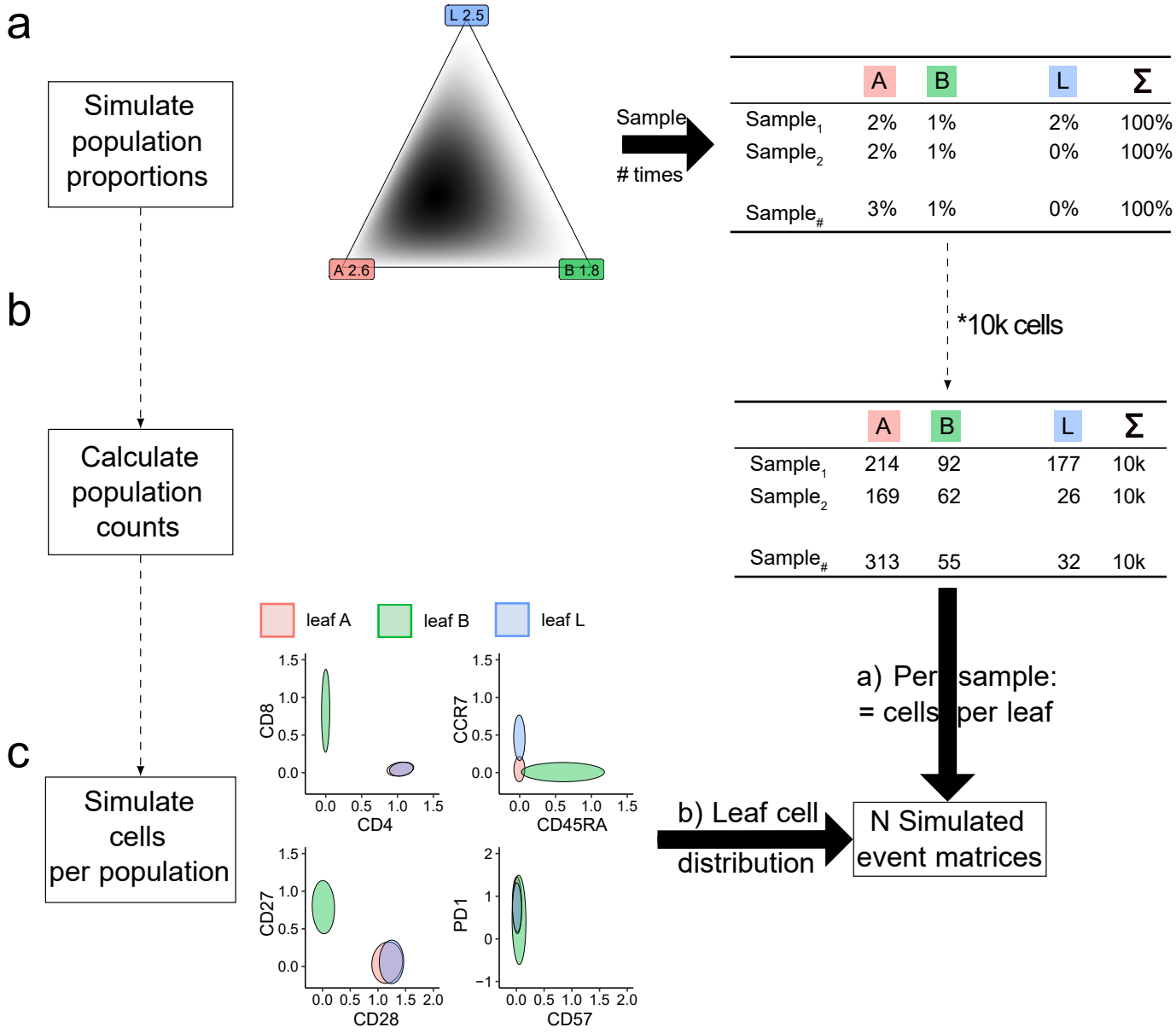
Supplementary Table 2

Subfigure	α_1	α_2	α_3	m_1	m_2	m_3	var_1	var_2	var_3	cov_{12}	cov_{13}	cov_{23}
a	1.00	1.00	1.00	0.33	0.33	0.33	0.06	0.06	0.06	-0.03	-0.03	-0.03
b	25.00	25.00	25.00	0.33	0.33	0.33	0.00	0.00	0.00	-0.00	-0.00	-0.00
c	1.00	1.00	10.00	0.08	0.08	0.83	0.01	0.01	0.01	-0.00	-0.01	-0.01
d	1.00	10.00	10.00	0.05	0.48	0.48	0.00	0.01	0.01	-0.00	-0.00	-0.01
e	1.00	1.00	0.10	0.48	0.48	0.05	0.08	0.08	0.01	-0.07	-0.01	-0.01
f	1.50	1.50	1.50	0.33	0.33	0.33	0.04	0.04	0.04	-0.02	-0.02	-0.02
g	0.10	0.10	0.10	0.33	0.33	0.33	0.17	0.17	0.17	-0.09	-0.09	-0.09
h	0.99	0.99	0.99	0.33	0.33	0.33	0.06	0.06	0.06	-0.03	-0.03	-0.03

Supplementary Table 2 *Parameters, means, variances and covariances for probability densities shown in Supplementary Figure 13.*

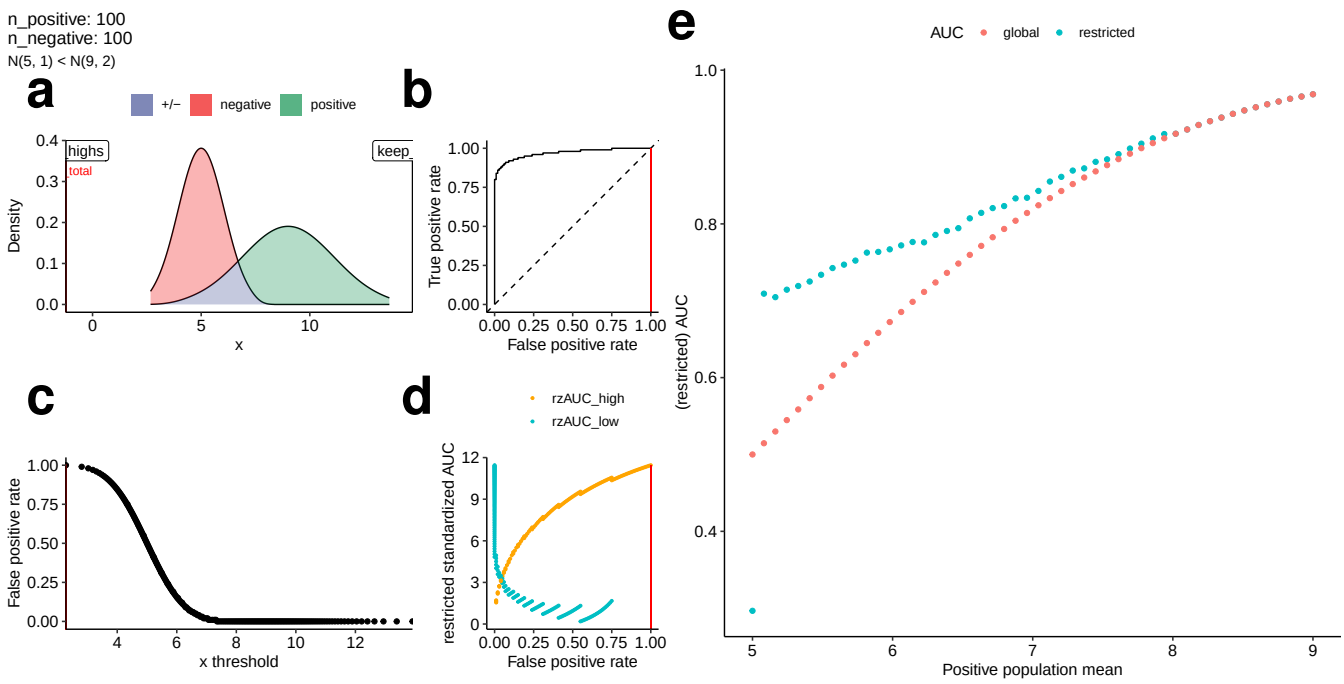
The table shows the parameters α_1, α_2 and α_3 from a Dirichlet distribution. Each row corresponds to the respective subfigure in Supplementary Figure 13. m_1, m_2 and m_3 are the means for each of the parameters ($m_i = E[p_i] = \frac{\alpha_i}{s}$ for $s = \sum_k \alpha_k$). var_1, var_2 and var_3 are the variances for each of the parameters ($var_i = Var[p_i] = \frac{\alpha_i(s-\alpha_i)}{s^2(s+1)}$). cov_{12}, cov_{13} and cov_{23} are the covariances for each of the parameters ($cov_{ij} = Cov[p_i, p_j] = \frac{-\alpha_i \alpha_j}{(s^2(s+1))}$).

Supplementary Figure 14



Supplementary Figure 14 *Procedure for generating realistic flow cytometry data.*
There are three steps to synthesizing new flow cytometry samples. **(a)** First, we generate cell population proportions for N new samples, where the percentage of cells in each gate are drawn from the previously established Dirichlet distribution. **(b)** Second, we calculate cell counts for all leaf gates by multiplying the simulated percentages by the desired overall number of cells; in this example, 10,000 total cells. **(c)** Third, we draw the required number of cells per leaf gate from its previously estimated multivariate normal distribution. This procedure is repeated for N newly synthetic samples. Thus, we obtain an event matrix of 10,000 cells for each of N samples.

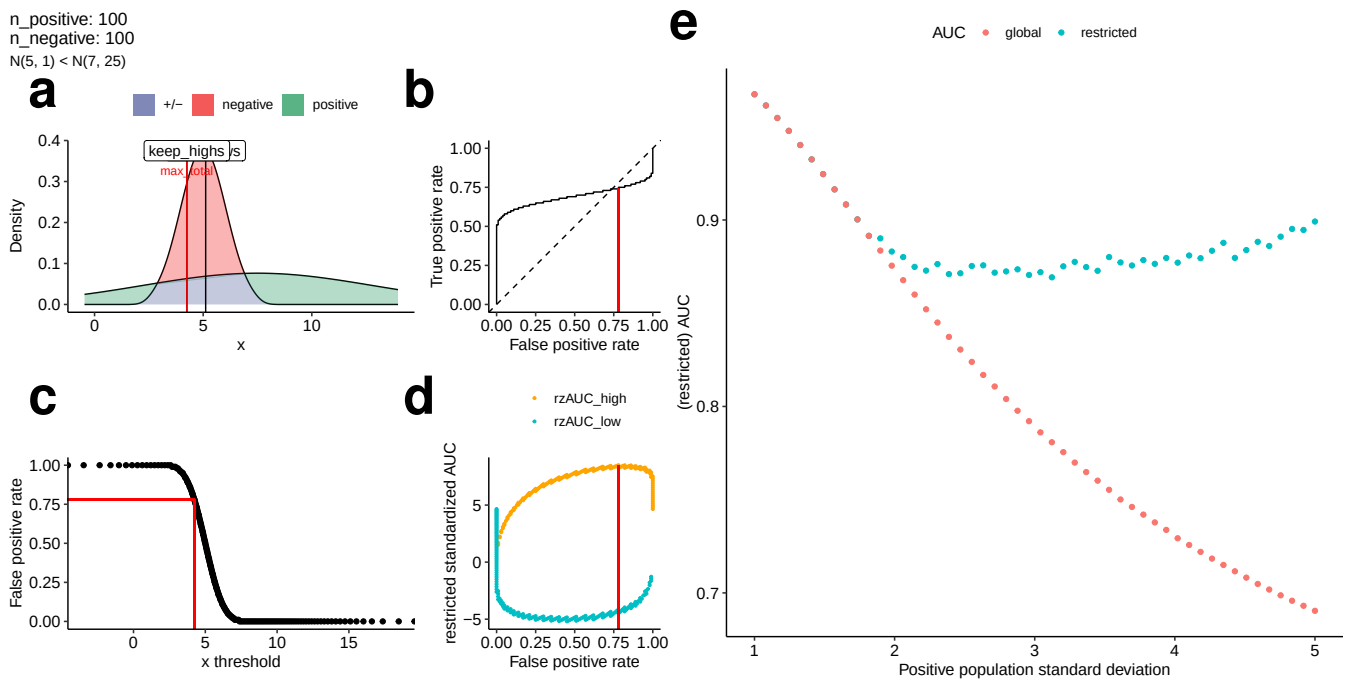
Supplementary video 1



Supplementary Video 1 *Restriction method applied to simulated distributions with varying mean of the positive population (GIF).*

We present simulated examples of marker normal distributions in two classes varying the positive population's mean. Positive and negative population are intended to represent sets of patients with different clinical outcomes with varying effect sizes. The distribution of values from positive (i.e. diseased) class are coloured green and values from negative (i.e. control) class are coloured red; the overlapping density areas are coloured in purple. For each example, we present the following 5 plots: **(a)** the positive and negative class densities; **(b)** the complete receiver operator characteristic (ROC) curve; **(c)** a plot of marker values against FPR; **(d)** a plot of rzAUC calculated for marker^{HIGH} (orange) and marker^{LOW} (blue) samples at all FPR values. **(e)** The restricted (blue) and unrestricted (red, "global") AUC against the mean of the positive population. In plots (a)-(d), red lines indicate the optimal restriction as a marker value or FPR value. The GIF shows the five plots when the mean of the positive population varies from $\mathcal{N}(5, 2)$ to $\mathcal{N}(9, 2)$ and the negative population remains $\mathcal{N}(5, 1)$. 100 samples are drawn from every population.

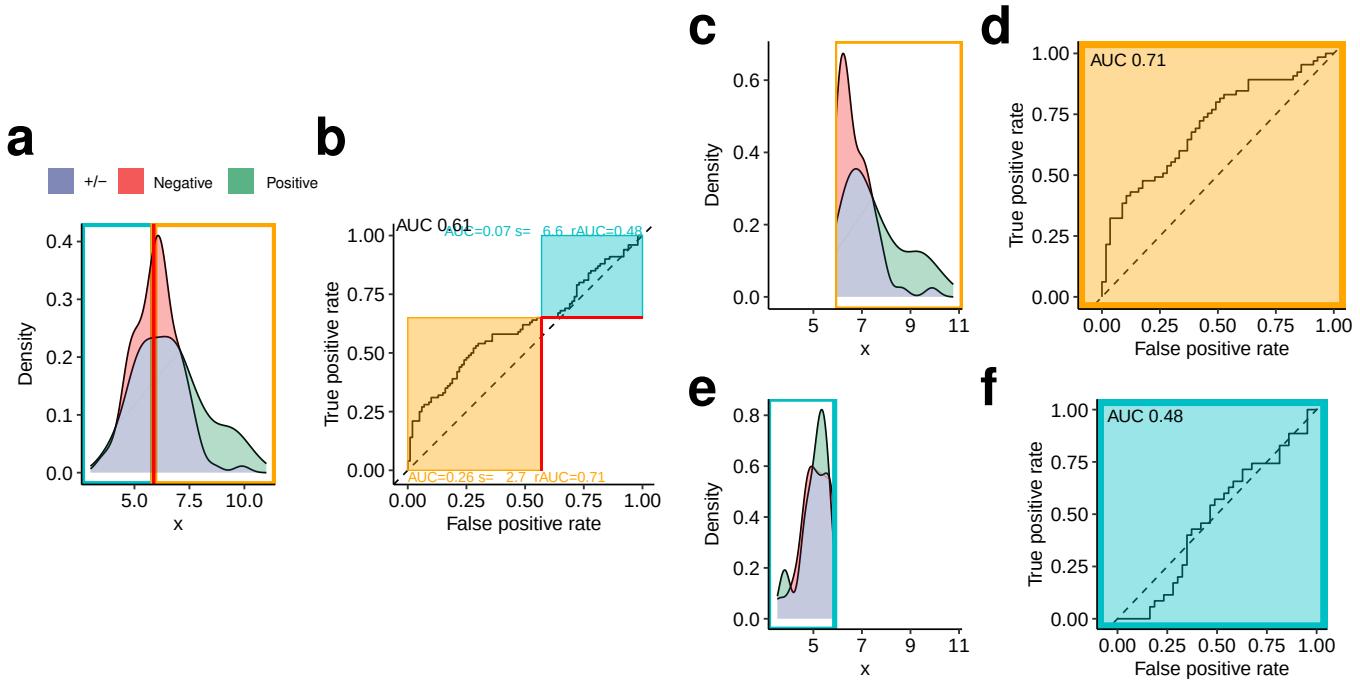
Supplementary video 2



Supplementary Video 2 *Restriction method applied to simulated distributions with varying variance of the positive population (GIF).*

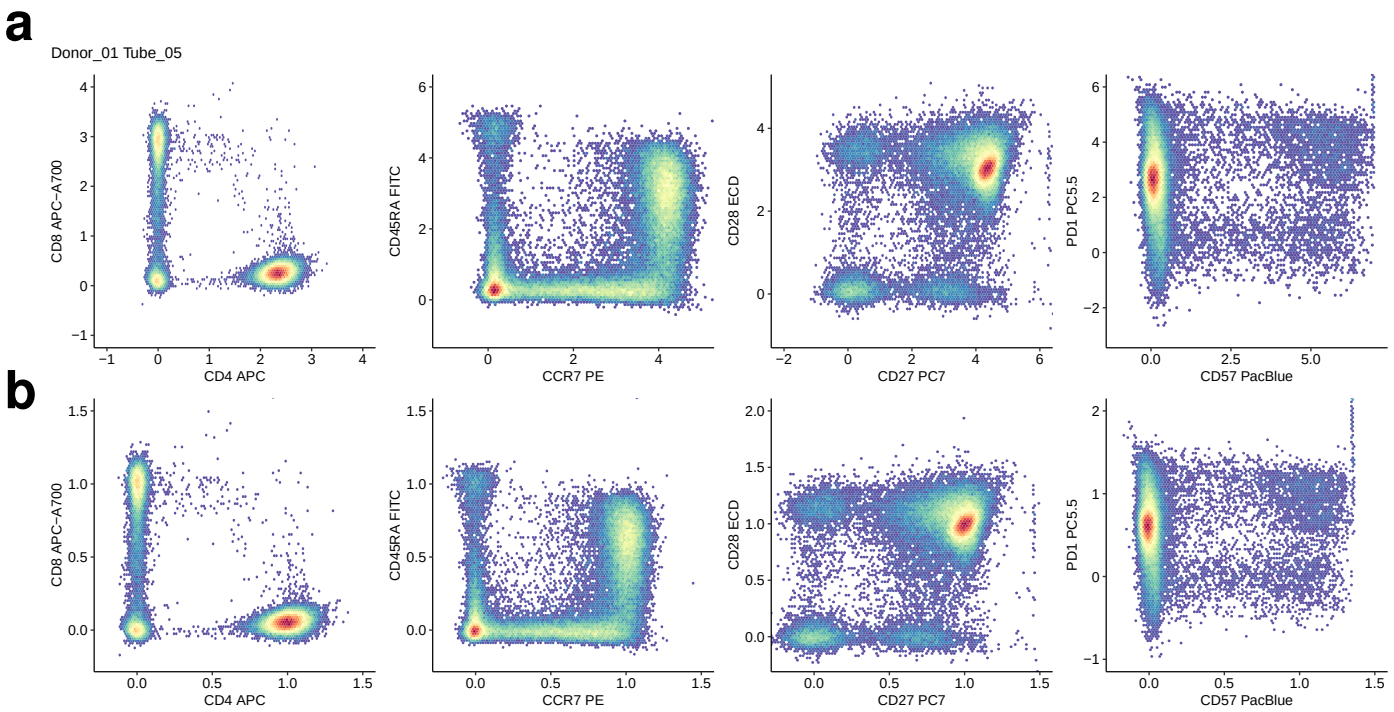
We present simulated examples of marker normal distributions in two classes varying the positive population's variance. Positive and negative population are intended to represent sets of patients with different clinical outcomes. The distribution of values from positive (i.e. diseased) class are coloured green and values from negative (i.e. control) class are coloured red; the overlapping density areas are coloured in purple. For each example, we present the following 5 plots: **(a)** the positive and negative class densities; **(b)** the complete receiver operator characteristic (ROC) curve; **(c)** a plot of marker values against FPR; **(d)** a plot of rzAUC calculated for $\text{marker}^{\text{HIGH}}$ (orange) and $\text{marker}^{\text{LOW}}$ (blue) samples at all FPR values. **(e)** The restricted (blue) and unrestricted (red, "global") AUC against the standard deviation of the positive population. In plots (a)-(d), red lines indicate the optimal restriction as a marker value or FPR value. The GIF shows the five plots when the variance of the positive population varies from $\mathcal{N}(7, 1)$ to $\mathcal{N}(7, 25)$ and the negative population remains $\mathcal{N}(5, 1)$. 100 samples are drawn from every population.

Supplementary video 3



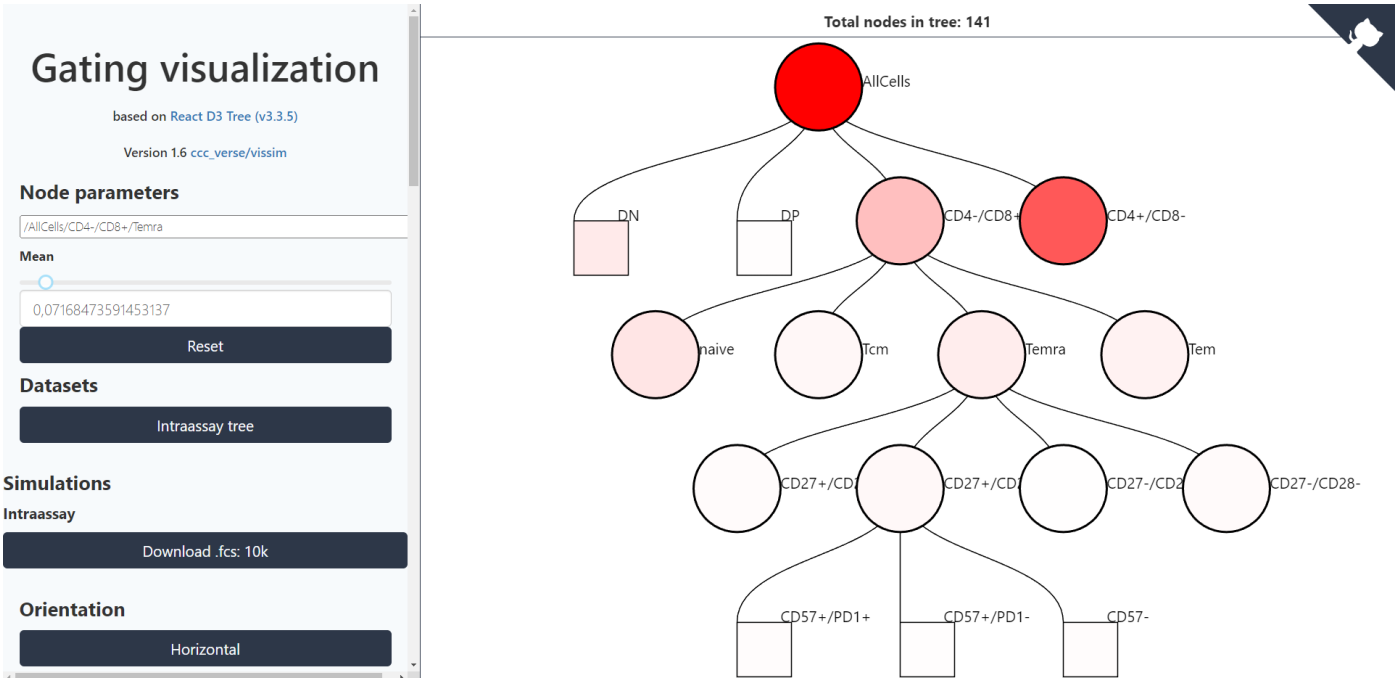
Supplementary Video 3 *Restricted AUC, correspondence of scaling factor and restriction (GIF).*
We present a simulated example of a marker distributions in two classes, which are intended to represent sets of patients with different clinical outcomes. We created a GIF for all possible restrictions and show the following 6 plots for each restriction. **(a)** The distribution of values from positive (i.e. diseased, n=100) class are coloured green and values from negative (i.e. control, n=100) class are coloured red; the overlapping density areas are coloured in purple. In this example, 20% of positive samples and 2% of negative samples were drawn from a population with elevated marker expression $\mathcal{N}(9,1)$ and all other samples were drawn from a population with unaltered marker expression $\mathcal{N}(6,1)$. Restriction of the dataset defines two subsets of samples - explicitly, marker$^{\text{HIGH}}$ (orange) and marker$^{\text{LOW}}$ (blue) samples. **(b)** The complete ROC curve marked at the current restriction (red lines). Restricting the parts of the ROC curve corresponding to marker$^{\text{HIGH}}$ (orange) or marker$^{\text{LOW}}$ (blue) samples gives us restricted ROC curves, for which a restricted AUC (rAUC) can be calculated. This is equivalent of rescaling the respective part of the ROC curve. In the plot, the complete AUC is shown (0.61). We show the rounded two-way partial area under the complete ROC curve for marker$^{\text{HIGH}}$ (orange) or marker$^{\text{LOW}}$ (blue) samples. Additionally, we show the rounded scaling factor s corresponding to reciprocal of the blue or orange rectangle area. The rAUC is then the two-way partial AUC multiplied with s. **(c)** The distribution of marker$^{\text{HIGH}}$ samples from positive and negative samples in the orange rectangle from (a) and (b). **(d)** ROC curve calculated only on marker$^{\text{HIGH}}$ samples. The complete AUC on only these marker$^{\text{HIGH}}$ samples is shown top-left in black and is identical to the orange rescaled rAUC from (b). **(e)** The distribution of marker$^{\text{LOW}}$ samples from positive and negative samples in the blue rectangle from (a) and (b). **(f)** ROC curve calculated only on marker$^{\text{LOW}}$ samples. The complete AUC on only these marker$^{\text{LOW}}$ samples is shown top-left in black and is identical to the blue rescaled rAUC from (b).

Supplementary video 4



Supplementary Video 4 *Asinh transformed and rescaled gating for all 48 samples (GIF).*
Following a conventional workflow, an experienced, blinded operator performed sample-wise manual recompensation of flow cytometry data using Kaluza software. Cell feature-wise asinh-transformations and rescaling were then applied. Here, we illustrate our data pre-processing with 48 samples, which represented 6 replicate stainings from 8 healthy donors. **(a)** Shows the gating of all cells from one donor's replicate (here donor 1, tube/replicate 5) after cell feature-wise asinh transformation. **(b)** Shows the gating of all cells from the same donor's replicate after cell feature-wise asinh transformation and rescaling.

Supplementary website 1



Supplementary Website 1 *vissim.gunthergl.com: Interactive gating tree to visualize our gating tree and to allow simulation of own flow cytometry datasets.*
A hierarchical gating tree used to classify peripheral blood leucocytes stained for T cell subset cell features. Within this hierarchy, intermediate gates (circles) define the cells that are categorized into end-nodes by leaf gates (rectangles). Typical frequencies of cells in all gates were estimated from 48 samples, which represented 6 replicate stainings from 8 healthy donors. Gates are coloured according to the percentage of all cells they encompass. We pre-processed the data by pre-filtering for $CD45^+$ cells and show further gates in the figure. We filter for $CD3^+$ T cells and subsequently split into quadrants according to CD4 at 0.19 and CD8 at 0.2 into double negatives (DN), double positives (DP), $CD4^+/CD8^-$ and $CD4^-/CD8^+$. Only $CD4^+/CD8^-$ and $CD4^-/CD8^+$ gates are split into quadrants according to CCR7 at 0.24 and CD45RA at 0.12 into TEM (-,-), TEMRA (-,+), TCM (+,-) and Tnaive (+,+). Every of those quadrants is again split into quadrants according to CD27 at 0.36 and CD28 at 0.30. Finally, every of those quadrants is split into $CD57^-$ (CD57<=0.23), $CD57^+/PD1^+$ (CD57>0.23 and PD1>0.52) and $CD57^+/PD1^-$ (CD57>0.23 and PD1<=0.52) cells. Therefore the interactive tree has 141 nodes. Selecting any node shows the node mean corresponding to its expected value (mean) of the Dirichlet distribution. The mean can be changed with the slider, therefore we can directly change the underlying distribution. Finally, an .fcs file containing 10,000 cells can be downloaded.

References

[1] Glehr G., Riquelme P., Yang Zhou J., Cordero L., Schilling H-L., Kapinsky M., Schlitt HJ., Geissler EK., Burkhardt R., Schmidt B., Haferkamp S., Hutchinson JA. & Kronenberg K. (2022). External validation of biomarkers for immune-related adverse events after immune checkpoint inhibition. In Front. Immunol. 13:1011040. doi: 10.3389/fimmu.2022.1011040

[2] Bamber, D. (1975). The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. In Journal of Mathematical Psychology (Vol. 12, Issue 4, pp. 387-415). Elsevier BV. https://doi.org/10.1016/0022-2496(75)90001-2

[3] Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. In Radiology (Vol. 143, Issue 1, pp. 29-36). Radiological Society of North America (RSNA). https://doi.org/10.1148/radiology.143.1.7063747

[4] Sullivan Pepe, M. (2004). The statistical evaluation of medical tests for classification and prediction. Oxford University Press.

[5] Dodd, L.E. (2001). Regression methods for areas and partial areas under the ROC curve. Ph.D. thesis, University of Washington

[6] Yang, H., Lu, K., Lyu, X., & Hu, F. (2017). Two-way partial AUC and its properties. In Statistical Methods in Medical Research (Vol. 28, Issue 1, pp. 184-195). SAGE Publications. https://doi.org/10.1177/0962280217718866

[7] Z. Yang, Q. Xu, S. Bao, Y. He, X. Cao and Q. Huang (2022). Optimizing Two-way Partial AUC with an End-to-end Framework". In IEEE Transactions on Pattern Analysis and Machine Intelligence. doi: 10.1109/T-PAMI.2022.3185311