Table 4. Specification of the 25 markers used to characterize cell sub-populations in our CyTOF experiments. Bead standards are embedded in each sample to allow Bead normalization. Each Bead contains the four heavy metal isotopes labeled by 1 in the third column.

Isotope	Marker	Beads
89Y	CD45	0
142Nd	CD19	0
143Nd	CD127	0
145Nd	CD4	0
146Nd	CD8a	0
147Sm	CD20	0
149Sm	CD25	0
151Eu	CD278	1
152Sm	TNFa	0
153Eu	Tim3	1
155Gd	CD27	0
156Gd	CD14	0
159Tb	CCR7	0
160Gd	CD28	0
161Dy	CD152	0
162Dy	FOXP3	0
164Dy	CD45RO	0
165Ho	INFg	1
166Er	CD223	0
167Er	GzB	0
170Er	CD3	0
172Yb	CD274	0
174Yb	HLADR	0
175Lu	PD1	1
209Bi	CD11b	0

1 Specification of Markers in CyTOF Experiments

Table 4 provides the information about the 25 markers used in the CyTOF experiments in Section 4.2.

2 Additional Results for CyTOF Calibration

Figure 8 shows the projection of the source and target samples onto the first two principal components of the target sample for the three additional source-target pairs not shown in Figure 1.

Figure 9 shows the empirical cumulative distribution functions of the next three markers before and after the calibration. In all cases, as well as on the remaining markers that are not shown here, the calibrated source curves are substantially closer to the target than the curves before calibration.

Figure 10 shows the projection of the CD8+ T-cell sub-population in the source and target data onto the first two principal components of the target sample for the three additional source-target pairs not shown in Figure 4.

Table 5 is analogous to Table 1, where the sample from day 2 was used as source and the sample from day 1 was used as target.

3 Indirect Calibration

In this section we demonstrate how MMD-ResNets can be used to calibrate a source distribution to a target distribution in an indirect manner, i.e., without training a net to learn this map directly, as in the previous experiments. For this experiment we use four of the CyTOF samples

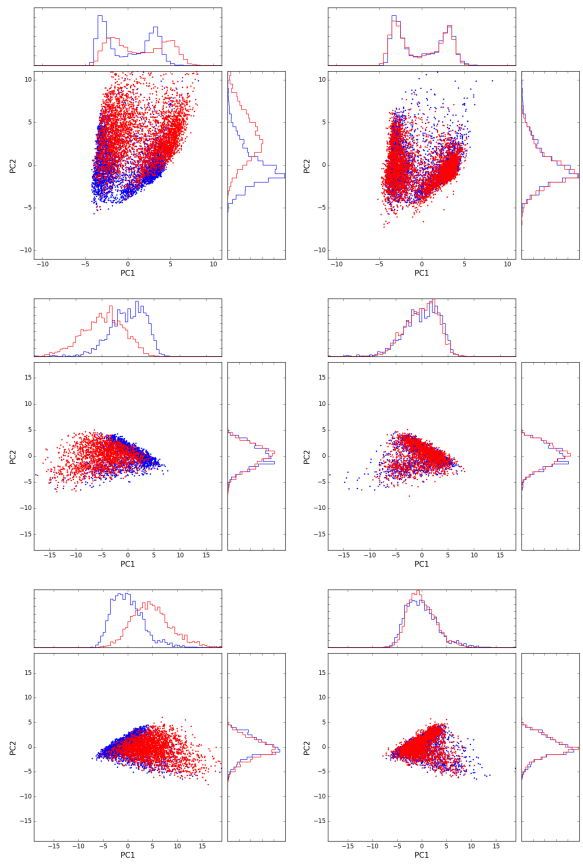

Fig. 8: Calibration of CyTOF data, for each of the three source-target pairs not shown in Figure 1. Projection of the source (red) and target (blue) samples on the first two principal components of the target data. Left: before calibration. Right: after calibration.

Table 5. CyTOF calibration experiment: MMD values between random batches of size 1000 from the source and target samples, before and after calibration on each of the four source-target pairs (patient1-baseline, patient2-baseline, patient1-treatment, patient2-treatment). In each pair, the sample from day 2 was used as source and the sample from day 1 was used as target. The MMD between two random batches of the target sample is provided as reference in the bottom row. The calibrated data is significantly closer in MMD to the target sample.

method \pair	pa.1 base.	pa.2 base.	pa.1 treat.	pa.2 treat.
no calibration	0.84	0.76	0.80	0.89
MLP calibration	0.22	0.16	0.26	0.30
ResNet calibration	0.18	0.16	0.17	0.19
MMD(target,target)	0.13	0.12	0.13	0.12

described in Section 4.2.1. i.e., samples from patients 1 and 2 at baseline condition, each measured on the instrument in day 1 and day 2. We use the shorthand notation p_1d_1 to refer to the sample of patient 1 measured in day 1 and similarly p_1d_2, p_2d_1, p_2d_2 to the other samples. In Section 4.2.3 we trained a MMD-ResNet (which we now denote by N_{p_1}) that maps p_1d_1 to p_1d_2 and a ResNet N_{p_2} which maps p_2d_1 to p_2d_2. In the following experiment we map p_1d_1 to p_1d_2 indirectly. The setup is as follows: In addition to the nets N_{p_1},N_{p_2} that were trained in Section 4.2.3, we train two additional MMD-ResNets, a ResNet N_{d_1}, mapping p_1d_1 to p_2d_1

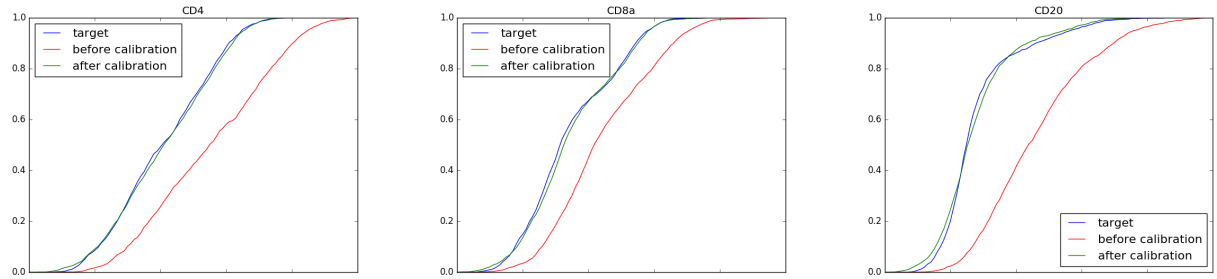

Fig. 9: A marginal perspective on the quality of calibration. Empirical cumulative distribution functions of the second three markers in the CyTOF calibration experiment. In each plot the blue, red and green curves corresponds to the target, source and calibrated source samples, respectively. In each marker the blue and green curves are substantially closer than the blue and red curves.

.

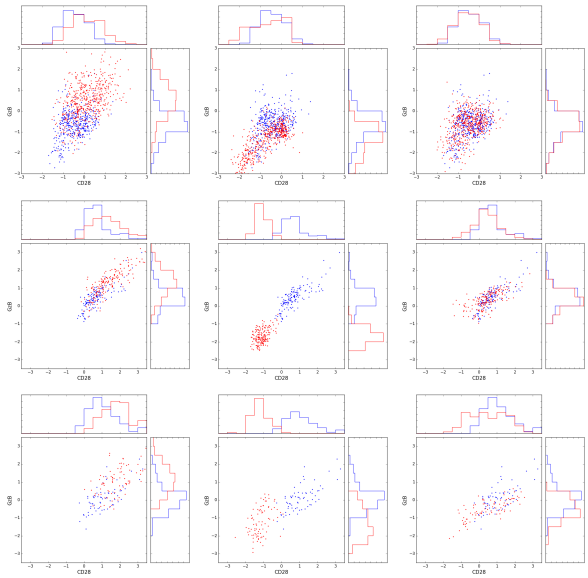

Fig. 10: Calibration of CD8+T-cells sub-population in the (CD28,GzB) plane, for each of the three source-target pairs not shown in Figure 4. In each row the left plot corresponds to before calibration, the right to calibration using ResNet, and the center to calibration using an identical net, without shortcut connections and initialized in a standard fashion.

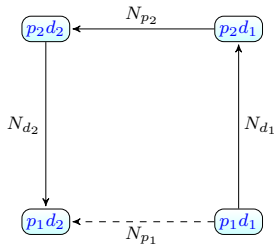

Fig. 11: Indirect calibration experiment scheme.

and a ResNet N_{d_2}, mapping p_2d_2 to p_1d_2. A scheme showing direct and indirect calibrations is shown in Figure 11. We then mapped p_1d_1 to p_1d_2 through N_{d_1}, followed by N_{p_2} and N_{d_2} (while adjusting the means and variances at each point, to account for the fact that each of these nets was trained on a standardized source sample), and compared the resulting calibration to the direct calibration obtained by applying N_{p_1} on p_1d_1. The results are presented in Figure 12. As we can see, the indirect calibration is only slightly less accurate than direct calibration, and removes much of

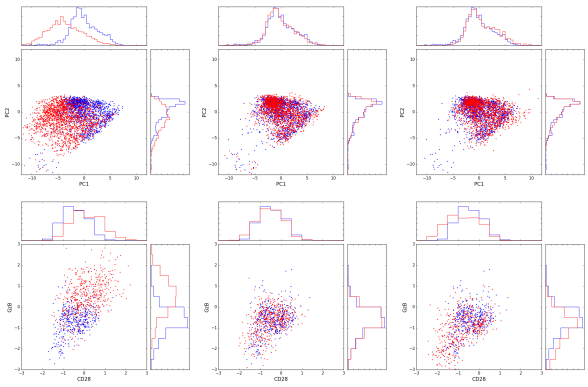

Fig. 12: Indirect calibration of CyTOF data. Left: before calibration. Center: direct calibration. Right: indirect calibration. Top row: whole sample, projected onto the subspace of the first two principal components. Bottom row: CD8 sub-population in the (CD28,GzB) plane.

the batch effect. MMD between the source and target values support this observation: while before calibration the MMD is 0.62, it is 0.19 after direct calibration and 0.20 after indirect calibration. The success of removing much of the batch effect via indirect calibration in the above experiment implies that the biological state of the patient p_1 was not distorted by much during the propagation through the nets N_{d_1} and N_{d_2}. This suggests that our MMD ResNets approach can be adapted for performing calibration in cases were replicates from a reference sample are measured in two batches and additional samples are measured only in one of the batches.

Suppose, for example, that in each day we run a CyTOF instrument to measure blood of several (different) individuals, and in addition we also measure in each of these runs a replicate of a reference blood sample. One can train a MMD-ResNet at each day t to calibrate the reference blood sample to its distribution at day 0. In addition, every replicate measured on day t can be calibrated using (a different) MMD-ResNet to the reference sample at day t. This way data from all days may be compared, by mapping all samples to coordinates of the reference sample at day 0.

4 Discussion

The problem of learning generative models has drawn much attention in the machine learning community recently. Evaluation of such models, however, is not always fully clear. Many recent works proposing generative models use Parzen window estimates for model evaluation. As Theis et al. (2015) nicely point out, evaluation of generative models using Parzen windows is problematic; in our context, for example, suppose that the net maps the source points to the centers of mass of the target sample. Such a map will have high Parzen likelihood estimates, while clearly not calibrating the data well. MMD, which takes also into account the internal structure of the calibrated source sample (term which is missing in Parzen estimates) might be more suitable for evaluation of the quality of the calibration.

In some of our experiments, which are not reported here, we found out that identifying cluster structure of the data might be a useful practice prior to applying MMD-ResNets in certain applications. For instance, when one uses CyTOF to characterize PBMC, the multi-marker cell distributions typically have separable clusters, corresponding to cell type sub-populations. While the relative proportion of different cell types in two replicate blood samples is expected to be invariant to the CyTOF machine, measuring these samples in two different runs in the same instrument or two different instruments often show noticeable differences between the cell type composition. When the proportions of corresponding clusters differ between the source and target distributions, we do not expect that a MMD-ResNet will account for that difference, as it computes a continuous map. In such cases, for example, it might be useful to use sub-sampling in order to match the relative proportions of each cell type between the source and the target samples.