Supplementary Note

Flow cytometry. All panels are detailed below, with antibody clones indicated in brackets (all reagents were obtained from BD Biosciences). Panels used to evaluate deep deconvolution (**Fig. 3a**) were configured using lyophilized reagent plates (Lyoplates, BD Biosciences), with the exception of reagents in parentheses, which were added as liquid antibodies.

Fig.	Tissue	Panel	*n*	FITC	PE	PerCP-Cy5.5	PE-Cy7	APC	APC-H7	V450	A700	Pac-Blue	APC-Cy7	Alexa-647
S3b	Tonsils	T/B cell	5	CD5 [L17F12]	-	-	-	CD19 [HIB19]	-	-	-	-	-	-
2h	Normal lung tissue	Leukocyte	11	CD4 [OKT4]	CD14 [HCD14]	CD19 [HIB19]	CD56 [HCD56]	CD8 [SK1]	-	-	CD45 [HI30]	-	-	-
2i,3c	FL lymph nodes	T/B cell	14	CD8 [SK1]	-	-	-	-	-	-	-	CD4 [RPA-T4]	CD20 [L27]	-
3a	PBMCs 1	T cell	20	(CD85j) [GHI/75]	(CD28) [L293]	CD4 [SK3]	CD45RA [HI100]	CD27 [L128]	CD8 [SK1]	CD3 [UCHT1]	-	-	-	-
3a	PBMCs 1	Activated T cell	20	(TCRgd) [11F2]	(PD-1) [EH12.1]	CD4 [SK3]	CD38 [HB7]	HLA-DR [L243]	CD8 [SK1]	CD3 [UCHT1]	-	-	-	-
3a	PBMCs 1	B cell	20	IgD [IA6-2]	CD24 [ML5]	CD19 [SJ25C1]	CD38 [HB7]	CD27 [L128]	CD20 [2H7]	CD3 [UCHT1]	-	-	-	-
3a	PBMCs 1	CXCR3+	20	CD16+56 [3G8/NCAM16.2]	CXCR3 [1C6/CXCR3]	CD4 [SK3]	CD33 [P67.6]	CD19 [SJ25C1]	CD8 [SK1]	CD3 [UCHT1]	-	-	-	-
3b	PBMCs 2	Treg	7	-	CD4 [SK3]	-	-	-	-	-	-	CD3 [UCHT1]	-	FOXP3 [236A/E7]

For **Supplementary Fig. 3b**, tonsil-derived cell suspensions were thawed, washed, counted, and subsequently stained with monoclonal antibodies (above table) to label B cells ($CD19^+$) and T cells ($CD5^+$), without stimulation. Each population was sorted using a FACSAria II instrument (BD Biosciences) to >95% purity for subsequent expression profiling.

For **Fig. 2h**, fresh normal lung tissue samples were cut into small pieces and dissociated into single cell suspensions by 45 min of Collagenase I (STEMCELL Technologies) digestion. Dissociated single cells were suspended at 1×10^7 per mL in staining buffer (HBSS with 2% heat-inactivated fetal calf serum). After 10 min of blocking with 10 μg/μL rat IgG, the cells were stained for at least 10 min with the antibodies indicated in the above table. After washing, stained cells were re-suspended in staining buffer with 1 μg/mL DAPI, and the following populations were enumerated using a FACSAria II instrument (BD Biosciences): total leukocytes ($CD45^+$), monocytes ($CD14^+$), CD8 T cells ($CD8^+$), CD4 T cells ($CD4^+$), NK cells ($CD56^+$), and B cells ($CD19^+$).

For **Figs. 2i** and **3c** (and **Supplementary Fig. 13**), diagnostic FL tumor cell suspensions were stained with monoclonal antibodies (above table) to label CD4 T cells ($CD4^+$), CD8 T cells ($CD8^+$), and B cells ($CD20^+$). Stained cells were detected on a FACSCalibur or an LSR II 3-laser cytometer (BD Biosciences).

For **Fig. 3a** (and **Supplementary Fig. 12a**), flow cytometry phenotyping was performed on PBMCs from healthy adults using lyophilized reagent plates (Lyoplates, BD Biosciences). The plates were configured with staining cocktails shown in the above table to enumerate the following cell subsets: naïve B cells ($CD3^-CD19^+CD20^+CD24^-CD38^+$), memory B cells ($CD3^-CD19^+CD20^+CD24^+CD38^-$), CD8 T cells ($CD3^+CD8^+$), naïve CD4 T cells ($CD3^+CD4^+CD45RA^+CD27^+$), memory CD4 T cells ($CD3^+CD4^+CD45RA^-$), gamma delta T cells ($TCRgd^+$), NK cells ($CXCR3^+CD16^+CD56^+$), and monocytes (identified by size via forward- and side-scatter properties). Staining was performed according to the published protocol for Lyoplates on an LSRII flow cytometer (BD Biosciences)[1]. Reagents in parentheses in the above table were added as liquid antibodies, and were not part of the Lyoplate *per se*.

Finally, for the enumeration of regulatory T cells (Tregs) in **Fig. 3b** (and **Supplementary Fig. 12b**), peripheral blood was obtained from six healthy adult males by venipuncture into K2EDTA vacutainers (BD Biosciences) and processed immediately. Whole blood was diluted two-fold with PBS and mononuclear cells (PBMCs) isolated using Ficoll-Paque Plus (GE Healthcare). PBMCs were washed twice with PBS, counted, and 1×10^6 cells per individual, along with 1×10^6 cells from viably preserved PBMCs obtained from patient 4 in **Supplementary Fig. 4c**, were stained with αCD3, and αCD4 (see table above). Cells were washed in PBS, resuspended in Fix/Perm Buffer (eBiosciences), and incubated on ice for 20 min. Cells were washed twice in Perm/Wash Buffer (eBiosciences), and stained with αFOXP3. Cells were washed once in Perm/Wash Buffer and data collected using an LSRFortessa flow cytometer (BD Biosciences). Tregs were defined as $CD3^+CD4^+FOXP3^+$ non-doublet cells, and enumerated as a fraction of all intact PBMCs.

Low frequency leukocyte subsets in PBMCs. Related to the main text, the following five leukocyte subsets had low median fractions in PBMCs as determined by flow cytometry (<5%): naïve and memory B cells, activated memory CD4 T cells, gamma delta T cells, and Tregs.

Supplementary Results

Enumeration of activated and resting memory CD4 T cells by flow cytometry. Characteristic changes in gene expression accompany the phenotypic transition from naïve ($CDR45RA^{+}CD45RO^{-}$) to memory ($CD45RO^{+}CDR45RA^{-}$) T cells. Two such genes were profiled in our activated T cell panel (**Supplementary Notes**, above): HLA-DR, a canonical T cell activation marker primarily expressed on memory CD4 T cells[2,3] (as opposed to naïve subsets), and CD38, another known activation marker predominantly expressed on naïve CD4 T cells[3,4]. While our activation T cell panel did not include CD45RA or CD45RO, we confirmed previous findings by analyzing data from a separate study (data not shown), in which PBMCs were profiled using a panel that included αCD3, αCD4, αCD45RA, αHLA-DR and αCD38. Among $CD3^{+}CD4^{+}$ cells in 6 healthy subjects, we confirmed a strong correlation between total $HLA\text{-}DR^{+}$ cells and $HLA\text{-}DR^{+}CD45RA^{-}$ (activated memory) cells ($R = 0.97$, $P = 0.001$; RMSE = 0.7%). Conversely, total $HLA\text{-}DR^{-}CD38^{+}$ counts were significantly correlated with $HLA\text{-}DR^{-}CD38^{+}CD45RA^{+}$ (naïve) cells ($R = 0.87$; $P = 0.001$; RMSE = 11.9%), suggesting that the $CD3^{+}CD4^{+}HLA\text{-}DR^{+}$ phenotype represents a reasonable surrogate for activated memory CD4 T cells in healthy adult PBMCs. Therefore, to compare flow cytometry data with activated and resting memory CD4 subsets (from LM22) in this study, we used counts of $CD3^{+}CD4^{+}HLA\text{-}DR^{+}$ cells to estimate levels of activated memory CD4 T cells, and subtracted these values from total memory CD4 T cells ($CD3^{+}CD4^{+}CD45RA^{-}$) to estimate resting memory CD4 T cells.

Analysis of feature selection. A key aspect distinguishing CIBERSORT from previous methods is the context-dependent selection of genes from the signature matrix. This procedure increases CIBERSORT's tolerance to noise and prevents overfitting[5] (Online Methods). Based on the underlying framework of SVR, we hypothesized that genes with highly variable expression across a signature matrix *S*, or between *S* and a mixture *M*, would be selected most frequently (e.g., see **Supplementary Fig. 1**, Online Methods). On the other hand, if feature selection is more heavily determined by the specific cell subsets from *S* present in *M*, then marker genes of a cell type missing from *M* might be discarded, potentially impacting performance on cell types closely related to the missing cell type. To evaluate these hypotheses, we created a simple spike series of two uncorrelated reference profiles from LM22, including pure samples of each (resting mast cells and CD8 T cells) (**Supplementary Fig. 9a**). We reasoned that if SVR excludes marker genes when corresponding cell types in *S* are absent from *M*, then this behavior would be most apparent in pure samples of uncorrelated cell types. Therefore, we first compared genes selected by SVR for 100% resting mast cells, but not 100% CD8 T cells, and vice

versa, and tested for differences in expression. Interestingly, genes uniquely selected for CD8 T cells were more highly expressed in CD8 T cells compared to resting mast cells, however the difference in magnitude was modest and the converse was not observed for resting mast cells (**Supplementary Fig. 9b**). This suggests a possible enrichment for marker genes based on mixture content, but the relationship was inconsistent. We then extended our analysis to signature matrix genes selected in common between the two cell types, and found many genes with highly variable expression, both between the two cell types and across LM22 (data not shown). Notable examples include *CD8A* (CD8 T cell-enriched) and *CPA3* (mast cell carboxypeptidase A3), *CLC*, and *TPSAB1* (MC enriched). Separately, when examining the spike series, highly expressed genes across all LM22 cell subsets were frequently selected despite the fact that analyzed mixtures contained only CD8 T cells and resting mast cells (**Supplementary Fig. 9c**). This is consistent with our former hypothesis, and suggests that signature matrix genes for a cell type present in *S* but absent from *M* are not necessarily discarded; rather, they are likely useful to CIBERSORT by bounding the regression (e.g., *CD8A* was chosen regardless of whether CD8 T cells were present, likely informing their absence; see **Supplementary Fig. 1**, Online Methods).

Importantly, our observation was reproducible with highly correlated cell subsets (data not shown). For example, when LM22 was applied to a pure sample of naïve CD4 T cells, *CD8A* was selected, despite not being expressed by naïve CD4 T cells. Further, the five genes most highly expressed in naïve CD4 T cells in LM22 (*IL7R*, *CD3D*, *TRAC*, *LTB*, *TRBC1*) were selected, despite being among the top 10 most highly expressed genes in CD8 T cells and resting CD4 memory T cells (**Supplementary Table 1**). Since these genes are highly variable in LM22 (e.g., generally low or absent in myeloid subsets) these data are also consistent with our former hypothesis, and suggest that cell subsets missing from the mixture are unlikely to adversely impact deconvolution of closely related subsets in the mixture.

Supplementary Discussion

Review of gene expression deconvolution methods. A variety of GEP deconvolution methods have been proposed, many of which represent a gene expression admixture and its components as a linear equation, **m** = **f** x **B**, where **f** denotes a vector consisting of the unknown fractions of each cell type and **B** is a GEP signature matrix. Previous groups have applied linear least squares regression (LLSR)[6] and more recently, non-negative least squares regression (NNLS)[7] and quadratic programming (QP)[8-10] to solve for **f** (**Supplementary Table 3**). While LLSR provides a maximum likelihood solution for **f** under

a normally distributed error model, the solution is approximate and does not enforce non-negativity constraints (i.e., cell population levels should never be <0)[8]. Though suboptimal, this issue can be adequately addressed in practice by setting negative coefficients to zero, followed by normalizing remaining coefficients to sum to 1. By contrast, NNLS and QP can explicitly incorporate non-negativity constraints, and QP can produce a globally optimal solution to **f** (in a least-squares sense) in practical time[8].

LLSR, NNLS, and QP generally display good performance if (i) **B** is well conditioned (i.e., its component cell types are highly distinct) and (ii) **B** is applied to a mixture sample whose components are largely known (e.g., mature immune populations in peripheral blood)[6,8]. However, these methods have major limitations for complex tissue analysis. First, because all data in **m** and **B** are used to solve for **f**, these approaches are not robust to noise or outliers, and not ideal for mixtures with considerable unknown content (i.e., cell types not incorporated into the signature matrix). Of note, unlike QP and LLSR, methods based on robust linear regression (RLR) perform a feature selection prior to regression (like CIBERSORT) and are therefore more resilient to outliers (e.g., Huber M-estimator regression using rlm in R), which may lead to better performance on complex tissues. Second, LLSR, NNLS, and QP require a well-conditioned signature matrix, and may exhibit decreased performance on cell types with highly similar GEPs, such as CD8 versus CD4 T cells, or naïve versus memory B cells (e.g., **Fig. 3d**). Such GEPs may exhibit multicollinearity, and can lead to a 'winner takes all' phenomenon in which a higher weight would be assigned to the cell type whose GEP is most concordant with the mixture, whereas slightly less correlated cell types would be disproportionately down-weighted.

Recently, several new GEP deconvolution methods (PSEA[11], DSA[10], MMAD[12], PERT[7]) were introduced that can impute relative fractions of cell subsets in **m** (**Supplementary Table 3**). Like previous approaches, PSEA and DSA solve the same system of linear equations for **f** using LLSR or QP, respectively (see above). Unlike other methods, PSEA and DSA require cell type specific marker genes as input, rather than signatures matrices, limiting their scope to fewer cell types. Moreover, because DSA derives expression levels in **B** from the input (i.e., marker genes and a set of mixtures **M**) rather than prior knowledge, mixtures with unknown content or noise may lead to reduced performance (**Supplementary Figs. 5**, **6**). The other two methods, MMAD and PERT, employ more complicated models. The former attempts to correct for normalization bias in expression data, while the latter estimates a multiplicative perturbation constant that acts equally on all mixtures, representing either biologically meaningful perturbations from reference profiles (e.g., cell culture effects) or noise. Both

use conjugate gradient descent to maximize the likelihood of **f** given their respective model assumptions. While theoretically more robust than simpler models, in practice, we found both methods to be similarly susceptible to complex mixtures with unknown content and noise (**Supplementary Figs. 5**, **6** and **Supplementary Table 4**).

Some GEP deconvolution methods rely on grouped samples instead of single samples[7,10,12,13], and some of these methods estimate the basis profiles at the same time as the cell type proportions[10] and/or estimate group cell type-specific differences[12,13]. In so doing, most of these methods estimate cell type proportions in each sample[7,10,12], but in a manner that depends upon the combinations of mixtures analyzed (data not shown). Such methods may be less reliable for single sample deconvolution, a desirable feature for personalized applications[14].

Finally, while previous GEP deconvolution methods have been validated using 'ground truth' mixtures with known cell type proportions[6,8,11], no general metric has been proposed to estimate the 'goodness of fit' between deconvolution results and new mixture samples. In our view, such a filter is critically important before GEP deconvolution is adopted more widely.

Supplementary References

1. Maecker, H.T. et al. *BMC Immunol.* **6**, 13 (2005).
2. Johannisson, A. & Festin, R. *Cytometry* **19**, 343-352 (1995).
3. Kestens, L. et al. *Clin. Exp. Immunol.* **95**, 436-441 (1994).
4. Prince, H.E., York, J. & Jensen, E.R. *Cell. Immunol.* **145**, 254-262 (1992).
5. Cherkassky, V. & Ma, Y. *Neural Netw* **17**, 113-126 (2004).
6. Abbas, A.R., Wolslegel, K., Seshasayee, D., Modrusan, Z. & Clark, H.F. *PLoS One* **4**, e6098 (2009).
7. Qiao, W. et al. *PLoS Comput. Biol.* **8**, e1002838 (2012).
8. Gong, T. et al. *PLoS One* **6**, e27156 (2011).
9. Gong, T. & Szustakowski, J.D. *Bioinformatics* **29**, 1083-1085 (2013).
10. Zhong, Y., Wan, Y.-W., Pang, K., Chow, L. & Liu, Z. *BMC Bioinformatics* **14**, 89 (2013).
11. Kuhn, A., Thu, D., Waldvogel, H.J., Faull, R.L.M. & Luthi-Carter, R. *Nat. Methods* **8**, 945-947 (2011).
12. Liebner, D.A., Huang, K. & Parvin, J.D. *Bioinformatics* **30**, 682-689 (2014).
13. Shen-Orr, S.S. et al. *Nat. Methods* **7**, 287-289 (2010).
14. Shen-Orr, S.S. & Gaujoux, R. *Curr. Opin. Immunol.* **25**, 571-578 (2013).