iScience, Volume *27*

Supplemental information

sNucConv: A bulk RNA-seq deconvolution method trained on single-nucleus RNA-seq data to estimate cell-type composition of human adipose tissues

Gil Sorek, Yulia Haim, Vered Chalifa-Caspi, Or Lazarescu, Maya Ziv-Agam, Tobias Hagemann, Pamela Arielle Nono Nankam, Matthias Blüher, Idit F. Liberty, Oleg Dukhno, Ivan Kukeev, Esti Yeger-Lotem, Assaf Rudich, and Liron Levin

Figure S1. Initial hVAT sample set snRNA-seq analysis, Related to Figure 2B,D

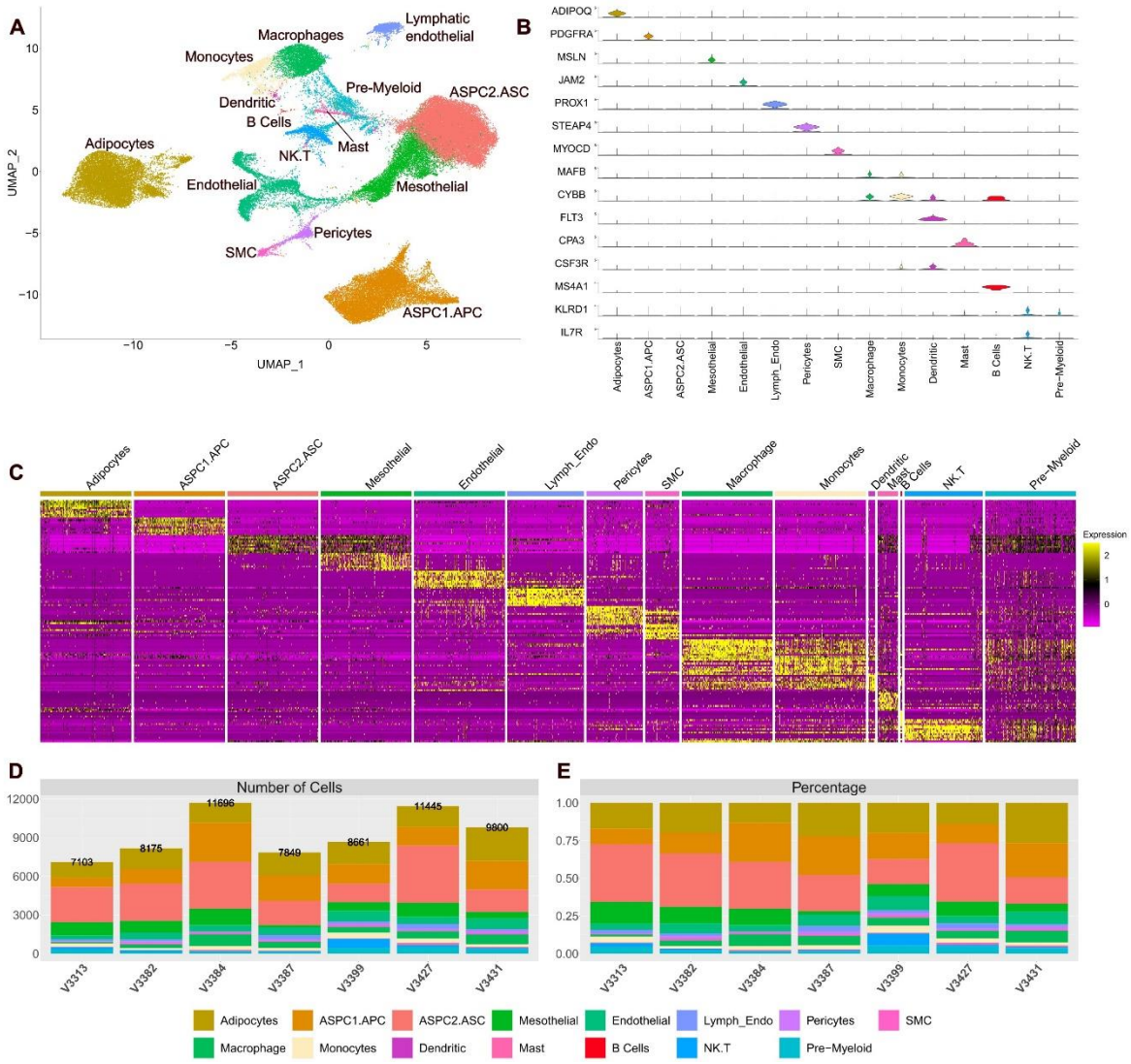

Figure S1: snRNA-seq of hVAT (initial sample set, n=7 samples, 64,729 nuclei)

UMAP (**A**) and violin plot (**B**) of single gene cell-type specific markers to identify 15 different cell-types in hVAT. **C.** Heatmap of displaying 10 preferentially expressed genes in each of the clusters (compared to all other nuclei). Nuclei numbers (**D**) and proportions (percentage, **E**) in each cluster per sample.

ASPC1.APC – Adipose stem and progenitor cells 1. Adipocyte progenitor cells; ASPC2.ASC - Adipose stem and progenitor cells 1. Adipose-derived stem cells; SMC – smooth muscle cells.

Figure S2. Initial hSAT sample set snRNA-seq analysis, Related to Figure 2C,E

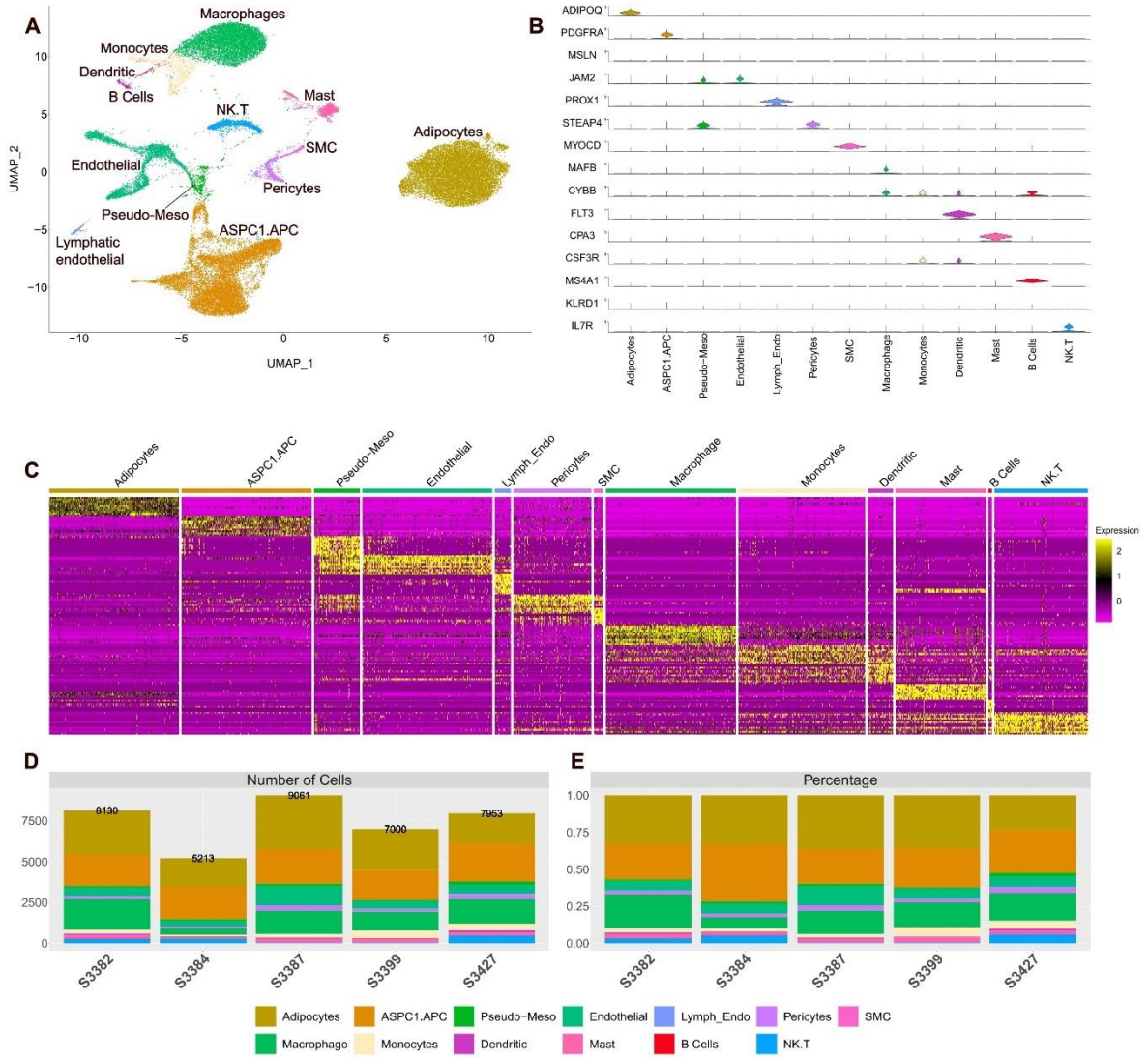

Figure S2: snRNA-seq of hSAT (initial sample set, n=5 samples, 37,357 nuclei)

UMAP (**A**) and violin plot (**B**) of single gene cell-type specific markers to identify 13 different cell-types in hSAT. **C.** Heatmap of displaying 10 preferentially expressed genes in each of the clusters (compared to all other nuclei). Nuclei numbers (**D**) and proportions (percentage, **E**) in each cluster per sample.

ASPC1.APC – Adipose stem and progenitor cells 1. Adipocyte progenitor cells; Adipose-derived stem cells; SMC – smooth muscle cells.

Figure S3. sNucConv performance, Related to Figure 4E,F

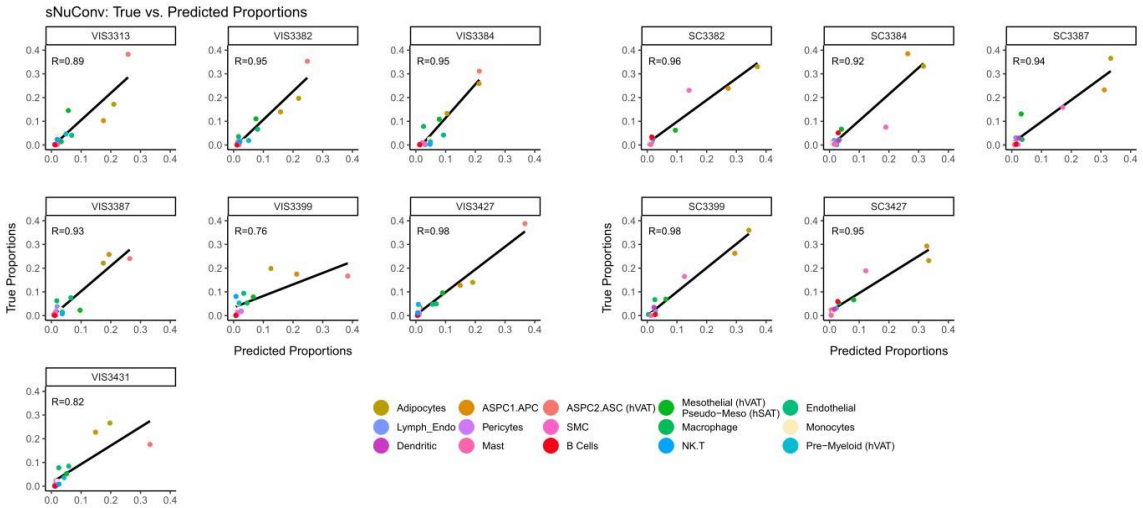

Figure S3: sNucConv estimated versus True (snRNA-seq -derived) cell-type proportion correlations: initial sample set (7 hVAT, 5 hSAT)

Graphs are scatter plots of correlations between sNucConv-estimated (applied unto bulk RNA-seq) versus the snRNA-seq -derived true cell-type proportions. Each graph represents the respective 15 and 13 cell types (in hVAT and hSAT) in each of the 12 initial sample set, using a leave-one-out approach.

Figure S4. sNucConv rarefaction analysis, Related to Figure 4E,F

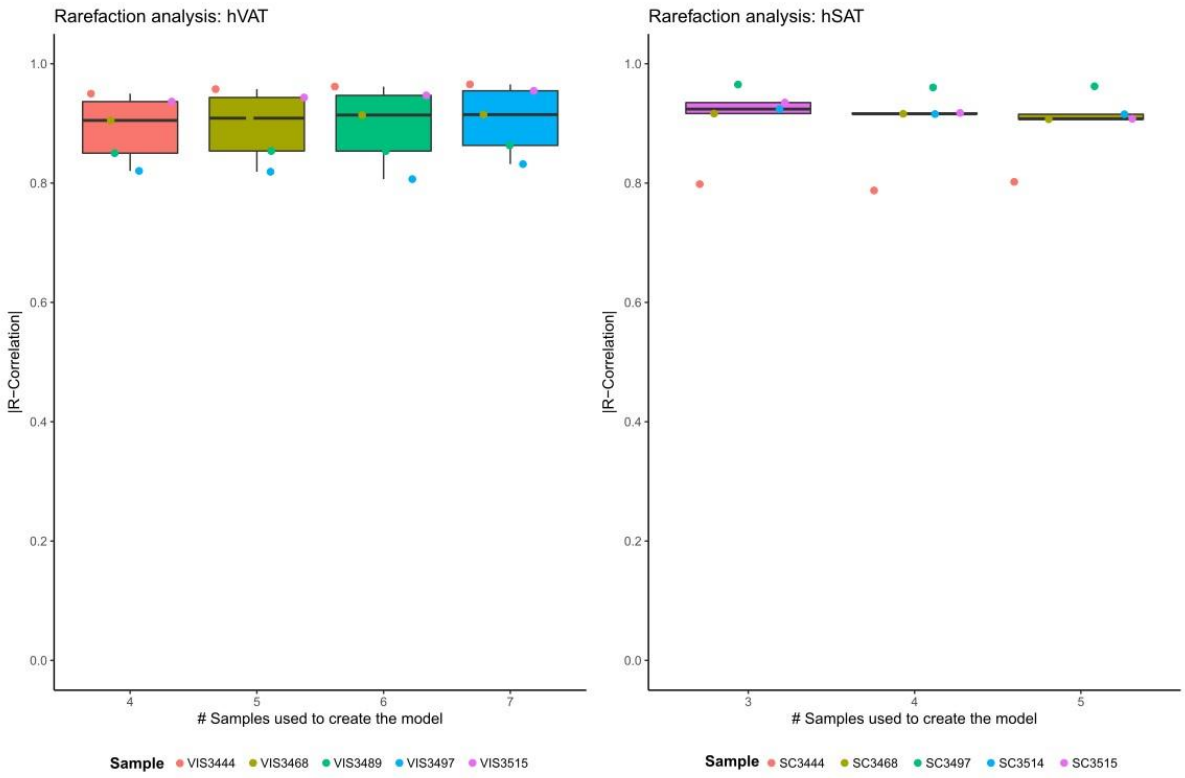

Figure S4: Rarefaction analysis of sNucConv trained on different number of samples from the initial sample set and assessed on the independent validation sample set.

sNucSeq was trained on 4,5,6 or 7 of the 7 hVAT initial sample set, or on 3,4 or 5 of the 5 hSAT (all possible combinations when not using all samples for training), and applied onto an independent sample set (validation sample set, 5 hVAT and 5 hSAT). Each point represent the mean R of correlation between the sNucConv-estimated and snRNA-seq -derived true cell-type proportions of all possible training models, per sample.

Figure S5. Doublets identification analysis, Related to method details “Integration and clustering”

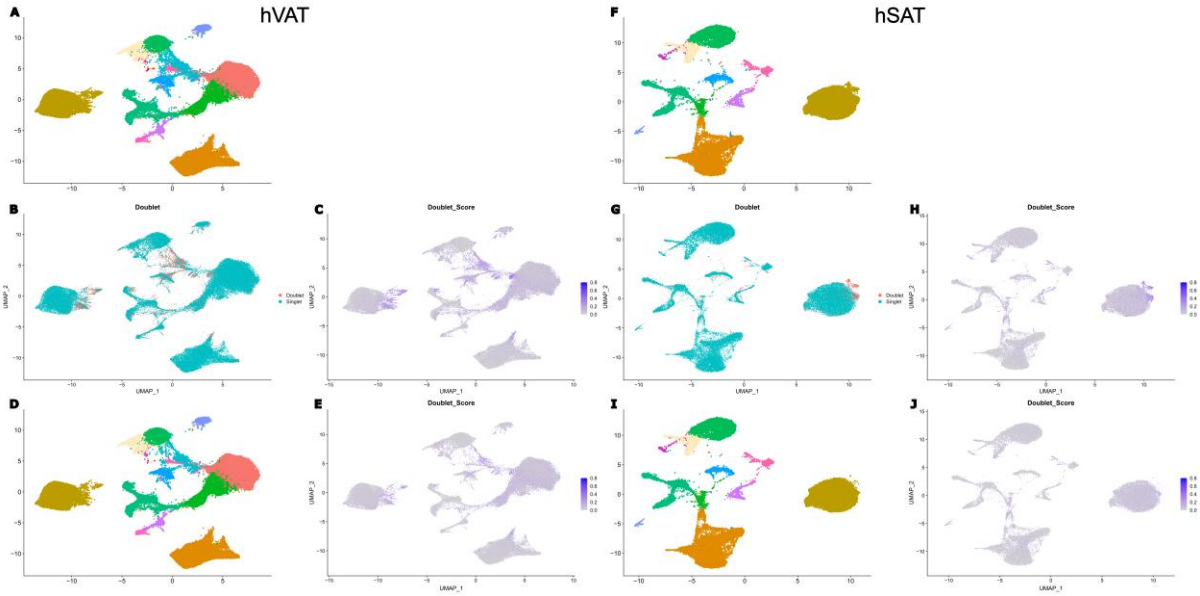

Figure S5: Secondary putative doublets identification analysis using Scrublet.

Five (5) UMAP panels are provided for each adipose tissue type (hVAT and hSAT). UMAPs showing clusters and cell-type annotations after DoubletFinder-based removal of putative doublets in hVAT and hSAT (**A., F., respectively**). UMAPs showing binary doublet/singlet identification as identified by Scrublet in hVAT and hSAT (**B., G., respectively**), and the respective UMAPs with doublet score of the entire nuclei as calculated by Scrublet in hVAT and hSAT (**C., H., respectively**). UMAPs showing clusters and cell-type annotations after DoubletFinder-based and additionally Scrublet-based removal of putative doublets in hVAT and hSAT (**D., I., respectively**), and the respective UMAPs with doublet score of the entire nuclei as calculated by Scrublet in hVAT and hSAT (**E., J., respectively**).

Table S1-A. Clinical and Biochemical Characteristics (N=7) – donors of the initial sample set, Related to experimental model and study participants details

	3313	**3382**	**3384**	**3387**	**3399**	**3427**	**3431**	**Mean** *(min-max)*
Age (y)	39	42	34	34	33	24	49	**36.4** *(24.0-49.0)*
Sex (male / female)	male	female	male	female	female	male	female	**3 / 4**
Body mass index (BMI, kg/m^2)	46.0	41.0	34.5	40.9	36.2	44.1	35.7	**39.8** *(34.5-46.0)*
Diastolic blood pressure (mm Hg)	79.0	87.0	61.0	80.0	88.0	95.0	70.0	**80.0** *(61.0-95.0)*
Systolic blood pressure (mm Hg)	130.0	123.0	116.0	117.0	133.0	132.0	128.0	**125.6** *(116.0-133.0)*
Triglycerides (mg/dL)	114.0	131.0	181.0	69.0	179.0	65.0	195.0	**133.4** *(65.0-195.0)*
Total cholesterol (mg/dL)	101.0	121.0	133.0	154.0	64.0	168.0	143.0	**126.3** *(64.0-168.0)*
Low-density lipoproteins (LDL, mg/dL)	56.0	89.0	119.0	99.0	28.0	102.0	56.0	**78.4** *(28-119)*
High-density lipoproteins (HDL, mg/dL)	38.0		48.0	42.0	38.0	54.0	48.0	**44.7** *(38.0-54.0)*
Fasting plasma glucose (FPG, mg/dL)	124.0	86.0	78.0	103.0	135.0	90.0	106.0	**103.1** *(78.0-135.0)*
Hemoglobin A1$_C$ (HbA1c, %)	5.4	5.4	5.5	5.8	6.2	5.8	6.7	**5.8** *(5.4-6.7)*
C-reactive protein (CRP)	1.47	0.46	0.64	0.97	0.38	0.45	0.90	**0.75** *(0.38-1.47)*

Table S1-B. Clinical and Biochemical Characteristics (N=6) – donors of the validation sample set, Related to experimental model and study participants details

	3444	**3468**	**3489**	**3497**	**3514**	**3515**	**Mean** *(min-max)*
Age (y)	31	25	24	55	30	30	**32.5** *(24.0-55.0)*
Sex (male / female)	female	female	female	male	female	female	**1 / 5**
Body mass index (BMI, kg/m^2)	49.0	35.8	55.0	52.0	48.0	47.0	**47.8** *(35.8-55.0)*
Diastolic blood pressure (mm Hg)	73.0	92.0	90.0	85.0	82.0	84.0	**84.4** *(73.0-92.0)*
Systolic blood pressure (mm Hg)	123.0	124.0	135.0	147.0	113.0	119.0	**126.8** *(113.0-147.0)*
Triglycerides (mg/dL)	88.0	173.0	144.0	152.0	159.0	115.0	**138.5** *(88.0-173.0)*
Total cholesterol (mg/dL)	118.0	165.0	119.0	156.0	196.0	145.0	**149.8** *(119.0-196.0)*
Low-density lipoproteins (LDL, mg/dL)	63.0	80.0	56.0	85.0	120.0	73.0	**79.5** *(56.0-120.0)*
High-density lipoproteins (HDL, mg/dL)	37.0	50.0	35.0	40.0	44.0	49.0	**42.5** *(35.0-50.0)*
Fasting plasma glucose (FPG, mg/dL)	108.0	99.0	112.0	114.0	79.0	117.0	**104.8** *(79.0-117.0)*
Hemoglobin $A1_C$ (HbA1c, %)	5.7	5.6	5.8	6.4	5.7	5.6	**5.8** *(5.6-6.4)*
C-reactive protein (CRP)	3.05	0.50	2.75	0.39	1.61	2.06	**1.72** *(0.5-3.05)*

Table S2-A. Cell Ranger parameters (raw counts, before QC) – Initial sample set, Related to methods details: "Quality control and mapping"

	Adipose tissue depot	Estimated number of nuclei	Mean reads per nucleus	Median genes per nucleus	Fraction reads in nuclei	Sequencing saturation
3313	vis	11,919	26,175	1,841	80.1%	54.7%
3382	sc	11,215	54,243	1,960	53.5%	61.4%
	vis	9,482	35,831	1,341	60.9%	74.0%
3384	sc	6,860	88,421	1,704	50.1%	76.2%
	vis	14,595	32,624	2,077	58.6%	32.2%
3387	sc	10,059	45,335	2,613	70.1%	48.0%
	vis	11,419	45,930	2,299	68.8%	44.1%
3399	sc	8,563	41,334	2,154	69.7%	40.5%
	vis	11,796	34,034	1,822	68.3%	48.5%
3427	sc	9,203	56,577	1,632	60.4%	65.2%
	vis	14,605	36,866	1,626	59.2%	56.7%
3431	vis	12,043	41,091	1,362	59.5%	70.9%
Mean	**sc**	**9,180**	**57,182**	**2,020**	**61%**	**58%**
	vis	**12,265**	**36,078**	**1,766**	**65%**	**54%**
min-max	*sc*	*6,860-11,215*	*41,334-88,421*	*1,632-2,613*	*50.1%*-70.1%	*40.5%-76.2%*
	vis	*9,482-14,595*	*26,175-45,930*	*1,341-2,299*	*58.6%-80.1%*	*32.2%-74.0%*

Table S2-B. Cell Ranger parameters (raw counts, before QC) – Validation sample set, Related to methods details: "Quality control and mapping"

	Adipose tissue depot	Estimated number of nuclei	Mean reads per nucleus	Median genes per nucleus	Fraction reads in nuclei	Sequencing saturation
3444	sc	9,768	51,632	1,790	62.4%	72.6%
	vis	13,012	37,987	1,718	65.0%	70.5%
3468	sc	7,343	62,712	1,950	74.3%	58.7%
	vis	9,482	50,581	2,178	69.7%	44.6%
3489	vis	10,328	46,101	1,604	65.5%	65.1%
3497	sc	4,617	75,947	1,671	63.0%	77.7%
	vis	6,278	44,343	1,675	66.4%	76.0%
3514	sc	9,698	54,851	1,789	70.1%	75.8%
3515	sc	9,255	69,244	1,646	60.7%	80.2%
	vis	7,479	60,108	1,790	59.7%	78.6%
Mean	**sc**	**8,136**	**62,877**	**1,769**	**66.1%**	**73.0%**
	vis	**9,315**	**46,624**	**1,793**	**65.2%**	**66.9%**
min-max	*sc*	*4,617-9,768*	*51,632-75,947*	*1,646-1,950*	*60.7%-74.3%*	*58.7%-80.2%*
	vis	*6,278-13,012*	*37,987-60,108*	*1,604-2,178*	*59.7%-69.7%*	*44.6%-78.6%*

Table S3. Studies developing deconvolution tools, Related to Discussion

Reference	Title	Tissue	scRNA-seq / snRNA-seq	n	# of cells/ nuclei	Deconvolution tool	Parallel bulk and sc/sn-RNA seq of same samples?	Performance assessment approach	Number of cell sub populations
Wang X et al., Nature Commu. 2019	Bulk tissue cell type deconvolution with multi-subject single-cell expression reference	Human pancreatic islets Mouse/rat kidney	scRNA-seq	*Human islets datasets:* n=31 scRNA-seq n=96 bulk-RNA seq *Mouse/rat kidney datasets:* n=7 scRNA-seq n=156 bulk-RNA seq	11,430 43,475	MuSiC	No indication No indication	Performance of MuSiC was compared to BSEQ-sc and CIBERSORT.	Human islets: 4 Mouse/rat kidney: 6
Menden K et al., *Sci. Adv.* 2020	Deep learning–based cell composition analysis from tissue expression profiles	Human PBMCs Human pancreas Human brain	scRNA-seq	*Human PBMCs datasets:* n=3 scRNA-seq n=24 bulk-RNA seq *Human pancreas datasets:* n=14 scRNA-seq n=18 bulk-RNA seq *Human brain datasets:* n=2 scRNA-seq n=390 bulk-RNA seq	26,219 12,083 27,881	Scaden	No indication No indication No indication	Performance of Scaden was compared to CSx, and MuSiC.	Human PBMCs: 6 Human pancreas:4 Human brain: 7
Dong M et al., *Briefings in Bioinformatics*. 2021	SCDC: bulk gene expression deconvolution by multiple single-cell RNA sequencing references	Mouse mammary glands	scRNA-seq	*Mouse mammary gland datasets:* n=2 scRNA-seq n=2 bulk-RNA seq	Not indicated	SCDC	Yes (n=2)	Performance of SCDC was compared to CSx, MuSiC, CIBERSORT and BSEQ-sc.	Mouse mammary glands: 5
Jew B et al. *Nature Commu*. 2020	Accurate estimation of cell composition in bulk expression through robust integration of single-cell information	Human SAT Human cortex	snRNA-seq snRNA-seq	*Human SAT datasets:* n=6 snRNA-seq n=106 bulk-RNA seq *Human cortex datasets:* n=8 snRNA-seq n=636 bulk-RNA seq	10,947 68,028	Bisque	Yes (n=6) Yes (n=8)	Performance of Bisque was compared to MuSiC, BSEQ-sc and CIBERSORT.	Human SAT: 5 Human cortex: 11
PRESENT STUDY: Sorek G et al. *iScience*	sNucConv: A bulk RNA-seq deconvolution method trained on single-nucleus RNA-seq data to estimate cell-type composition of human subcutaneous and visceral adipose tissues	Human SAT Human VAT	snRNA-seq snRNA-seq	*Human SAT datasets:* **n=10** snRNA-seq **n=10** bulk-RNA seq *Human VAT datasets:* **n=12** snRNA-seq **n=12** bulk-RNA seq	86,581 132,438	sNucConv	Yes (n=10) (5 "initial" set + 5 validation set) Yes (n=12) (7 "initial" set + 5 validation set)	Performance of sNuConv was compared to MuSiC, Scaden, SCDC and CIBERSORTx(1).	Human SAT: 13 Human VAT: 15