1 Supplementary notes

Proof 1 *(Theorem 1)* $V^{(m)}V^{(m)^T}$ *is the projection matrix that projects an arbitrary vector to the principal component space of group* m*. From the definition we have* $\cos^2(\delta^{(m)}) = b^T V^{(m)}V^{(m)^T}b$*. Therefore* $\sum_{m=1}^{M}\cos^2(\delta^{(m)}) = b^T(\sum_{m=1}^{M}V^{(m)}V^{(m)^T})b$ *is maximized by the eigenvector corresponding to the largest eigenvalue* λ_1 *of* $\sum_{m=1}^{M}V^{(m)}V^{(m)^T}$*.*

Theorem 2 *We already showed in Theorem 1 that* $\sum_{m=1}^{M}cos^2\delta^{(m)}$ *is maximized by* $\mathbf{b}_1$ *(eigenvector of* $V^{(m)}V^{(m)^T}$*). Repeating this procedure* k *times by using the eigenvectors corresponding to the first* k *largest eigenvalues will lead to the subspace of dimension* k *which resembles all* M *group closely. E.g. the eigenvector* $\mathbf{b}_2$ *corresponding to the second largest eigenvalue will give the direction, which is orthogonal to* $\mathbf{b}_1$ *and* $\sum_{m=1}^{M}cos^2\delta^{(m)}$ *obtains its next largest value.*

Proof 2 *(Theorem 2) Since* $\sum_{m=1}^{M}\cos^2(\delta^{(m)}) = b^T(\sum_{m=1}^{M}V^{(m)}V^{(m)^T})b$*, within the subspace* $\mathbb{R}^p/\{\mathbf{b}_1\}$*,* $\sum_{m=1}^{M}\cos^2(\delta^{(m)})$ *will be maximized by eigenvector* $\mathbf{b}_2$ *corresponding to the second largest eigenvalue* λ_2 *of* $\sum_{m=1}^{M}V^{(m)}V^{(m)^T}$*. This procedure can be repeated* k *times and the resulting* $\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_k$*, will best resemble all* M *groups of principal components as closely as possible.*

2 Simulation studies under various scenarios

Simulation studies via comprehensive experimental scenarios are taken into account that can capture most real applications: (1) different strength of dependence (2) full dependence (3) different sample size for different studies. In what follows, we spell down technical details of simulating omics data, from which we differently created the underlying true common covariance matrix Σ*. Table S7 encapsulates all numerical results of each simulation design. It is clear to say that SV and SSC consistently outperform JIVE, standard PCA and pooled PCA over the experiment scenarios.*

2.1 Different strength in block diagonal structures

$\Sigma = e_1^T \lambda_1 e_1 + e_2^T \lambda_2 e_2 + \Theta_\rho$, $\Theta_\rho = \{\theta_{ij}\}$ *and* $\theta_{ij} = \rho\ (= 0.5)$ *if* $1 \leq i, j \leq 50$ *or* $51 \leq i, j \leq 100$, *otherwise* $\theta_{ij} = 0$. *This configuration serves to impose gene correlation structures to* Σ. *We simulated covariance matrix* $\Sigma^{(m)}$ *for the* m^{th} *study* ($1 \leq m \leq M$), *where* $\Sigma^{(m)} = \Sigma + E^{*(m)}$, $E^{*(m)} = E^{(m)^T} \cdot E^{(m)}$, $E^{(m)} = \left(\epsilon_1^{(m)}, \ldots, \epsilon_{200}^{(m)}\right)$, $\epsilon_i^{(m)} \sim MVN_p(\boldsymbol{0}, W)$, $W = C \cdot I$, $C \in \{0.1, 0.5, 1\}$ *and* C *functions as the noise level and* $I_{p\times p}$ *is an identity matrix. We generated* M *simulated datasets of 20 samples and 200 features,* $X^{(m)} = (x_1^{(m)}, \cdots, x_{200}^{(m)}) \sim MVN_{200}(\boldsymbol{0}, \Sigma^{(m)})$ *for* $1 \leq m \leq M$ *and* $1 \leq M \leq 10$.

2.2 Full dependence

$\Sigma = e_1^T \lambda e_1 + e_2^T \lambda e_2 + I + \Theta_\rho$, $\Theta_\rho = \{\theta_{ij}\}$, *and* $\theta_{ij} = \rho\ (= 0.5)$ *for* $1 \leq i, j \leq 200$. I *is a diagonal matrix to constantly keep the positive definite property. The remaining designs are identical to (2.1)*

2.3 Different sample size

This simulation design is equivalent to (2.1) but just differs in sample size. To figure out the sheer effect of sample size, we just augment sample size adding up to 40.

2.4 Different measurement scales

It is worth of considering the effects of different measurement scales as various microarray platforms are employed in practice. To this end, we simulated datasets with varied mean, say, $X^{(m)} =$

$(x_1^{(m)}, \cdots, x_{200}^{(m)}) \sim MVN_{200}(\mu^*, \Sigma^{(m)})$ *for* $1 \leq m \leq M$ *and* $1 \leq M \leq 10$, *where* $\mu^* = (\underbrace{\mu, \mu, \ldots, \mu}_{200})$, $\mu \sim unif(0, 10)$. *The remaining simulation settings are equivalent to (2.1).*

3 Flury's likelihood ratio statistic

In this section, we evaluate if the likelihood ratio test is practically applicable to high-dimension omics data. Below we outline our simulation setting:

***Step 1 (True eigen-space)**: We considered two-dimensional underlying true eigen-space spanned by* $E = (e_1^T, e_2^T)$ *and* $\lambda = (\lambda_1, \lambda_2)$ *be the corresponding true eigenvalues, where* $e_1 = (1, 0, \ldots, 0) \in R^{1\times p}$ *and* $e_2 = (0, 1, \ldots, 0) \in \mathbb{R}^{1\times p}$, $p = 10, 50$ *and* 100, $\lambda_1 = 2$ *and* $\lambda_2 = 1$.

***Step 2 (Simulate datasets)**: By multiplying the true eigenvectors and eigenvalues, we created the underlying true common covariance matrix* Σ, *where* $\Sigma = e_1^T\lambda_1 e_1 + e_2^T\lambda_2 e_2$. *This configuration serves to impose gene correlation structures to* Σ. *We simulated covariance matrix* $\Sigma^{(m)}$ *for the* m^{th} *study* $(1 \leq m \leq M)$, *where* $\Sigma^{(m)} = \Sigma + E^{*(m)}$, $E^{*(m)} = E^{(m)^T} \cdot E^{(m)}$, $E^{(m)} = \left(\epsilon_1^{(m)}, \ldots, \epsilon_p^{(m)}\right)$, $\epsilon_i^{(m)} \sim MVN_p(\boldsymbol{0}, W)$, $W = C \cdot I$, $C = 10^{-4}$ *and* C *functions as the noise level and* $I_{p\times p}$ *is an identity matrix. We generated* M *simulated datasets of 1000 samples and 10, 50 and 100 features,* $X^{(m)} = (x_1^{(m)}, \cdots, x_p^{(m)}) \sim MVN_p(\boldsymbol{0}, \Sigma^{(m)})$ *for* $1 \leq m \leq 2$.

In what follows, we revisit the likelihood test statistic (Flury, 1984)

$$\sum_{m=1}^{M} n^{(m)} log\frac{|\hat{\Sigma}^{(m)}|}{|S^{(m)}|} \sim \chi^2_{(m-1)p(p-1),} \tag{1}$$

where $S^{(m)}$ *is the sample covariance matrix of study* m, $\hat{\Sigma}^{(m)} = \hat{B}\Lambda^{(m)}\hat{B}^T$, $\hat{B}$ *and* $\Lambda^{(m)}$ *are the estimated common eigenvector and eigenvalues matrix, respectively. We repeat this simulation 1,000 times to access the true type I error rate at a nominal level of* $\alpha = 0.05$. *As feature size increases* $(p = 10, 50, 100)$ *with sample size fixed* $(n = 1,000)$, *type I error is found notably increasing as 0.048, 0.255 and 1, respectively. This implies that the Flury's likelihood ratio test (1) is only suited to small-p and large-n experiments. This is expected since the likelihood ratio test is an asymptotic test which requires large number of samples. For high dimensional setting, Flury's likelihood ratio test gets even worse.*

Table S1: Sample size for multi-tissue microarray studies using metabolism related knockout mice, These microarray date are from Li et al. (2011). They are of the same platform and each tissue has 14,495 genes. There are four tissues to be meta-analyzed comparing wild type (WT), very long chain acyl-CoA dehydrogenase deficient (VLCAD-deficient) and long chain acyl-CoA dehydrogenase deficient (LCAD-deficient).

Tissue	wild type	VLCAD-deficient	LCAD-deficient
brown	4	4	4
heart	3	4	3
liver	4	4	4
skeleton	3	3	3

Table S2: The summary of four prostate cancer data.

Author	Year	Platform	Sample Size	Source
Lapointe et al.	2004	cDNA	103	GSE3933
Tomlins et al.	2006	cDNA	57	GSE6099
Varambally et al.	2005	HG-U133 Plus 2	13	GSE3325
Yu et al.	2004	HG-U95Av2	146	GSE6919

Table S3: The summary of six TCGA methylation data.

Type	Platform	# of genes	Sample Size	Source
BRCA	HumanMethylation27	13,311	350	The Cancer Genome Atlas
COAD	HumanMethylation27	13,169	215	The Cancer Genome Atlas
KIRC	HumanMethylation27	12,606	427	The Cancer Genome Atlas
LUAD	HumanMethylation27	12,709	157	The Cancer Genome Atlas
READ	HumanMethylation27	13,295	84	The Cancer Genome Atlas
STAD	HumanMethylation27	13,196	114	The Cancer Genome Atlas

Table S4: The four proposed methods of sparse MetaPCAs for variable selection

Method 1: SSC + PMD

Estimate Meta-sparse eigenvectors (SSC) by the sparse PCA method (PMD),

$(U^*, B^*) = \text{argmax}_{U,B} U^T T^{SSC} B$ subject to $\|B\|_2^2 \leq 1, \|B\|_1 \leq \lambda$ and $\|U\|_2^2 \leq 1$,

where $B^* = (\beta_1^*, \cdots, \beta_K^*) \in R^{p \times K}$

Method 2: SSC + eNet

Estimate Meta-sparse eigenvectors (SSC) by the sparse PCA method (eNet),

$(A^*, B^*) = \text{argmax}_{A,B} \sum_{i=1}^{p} \|t_i - AB^T t_i\|^2 + \sum_{j=1}^{K} \|\beta_j\|^2 + \sum_{j=1}^{K} \lambda_j \|\beta_j\|_1$ subject to $A^T A = I$,

where $B^* = (\beta_1^*, \cdots, \beta_K^*) \in R^{p \times K}$ and t_i is the i^{th} column vector of $T^{SSC} (1 \leq i \leq p)$.

Method 3: SV + PMD

Estimate Meta-sparse eigenvectors (SV) by the sparse PCA method (PMD)

$(U^*, B^*) = \text{argmax}_{U,B} U^T T^{SV} B$ subject to $\|B\|_2^2 \leq 1, \|B\|_1 \leq \lambda$ and $\|U\|_2^2 \leq 1$,

where $B^* = (\beta_1^*, \cdots, \beta_K^*) \in R^{p \times K}$.

Method 4: SV + eNet

Estimate Meta-sparse eigenvectors (SV) by the sparse PCA method (eNet),

$(A^*, B^*) = \text{argmax}_{A,B} \sum_{i=1}^{p} \|t_i - AB^T t_i\|^2 + \sum_{j=1}^{K} \|\beta_j\|^2 + \sum_{j=1}^{K} \lambda_j \|\beta_j\|_1$ subject to $A^T A = I$,

where $B^* = (\beta_1^*, \cdots, \beta_K^*) \in R^{p \times K}$ and t_i is the i^{th} column vector of T^{SV} $(1 \leq i \leq p)$.

Table S5: Fisher discriminant scores of PC projections (mouse metabolism data; Brown denotes Brown fat).

	Brown	Heart	Liver	Ske	Average
Brown	8.64	12.60	7.75	8.15	9.28
Heart	16.65	24.43	15.28	10.91	16.82
Liver	3.83	5.48	2.19	5.23	4.18
Ske	15.51	16.91	12.93	20.93	16.57
SSC	8.28	15.05	8.40	8.93	10.17
JIVE	3.59	5.83	3.75	3.35	4.13
SSC+PMD	19.11	29.17	22.90	22.68	23.47

Table S6: Fisher discriminant scores of PC projections (TCGA pan-cancer data; Class lables: Tumor, Normal, Male and Female).

	BRCA	COAD	KIRC	LUAD	READ	STAD	Average
BRCA	18.16	22.22	20.73	12.17	15.04	8.17	16.08
COAD	20.50	25.50	28.23	13.87	17.50	10.70	19.38
KIRC	22.59	29.13	32.70	16.25	20.33	13.78	22.46
LUAD	21.81	25.30	27.64	14.47	17.03	11.09	19.55
READ	20.27	21.29	18.43	11.06	15.35	7.02	15.57
STAD	21.84	26.17	29.34	14.89	17.40	11.98	20.27
SSC	24.93	21.02	16.88	12.52	13.12	7.94	16.07
JIVE	19.69	20.15	18.50	10.68	12.77	8.90	15.11
SSC+PMD	16.96	29.66	27.12	14.72	20.34	13.98	20.46

Table S7: Evaluation measures of simulated data sets for all experiment scenarios.

# of studies	2	3	4	5	6	7	8	9	10
A. different strength in block diagonal structures (implemented in the main manuscript)									
	experiment parameters (C=1, rho=0.5)								
SV	0.4430	0.6740	0.8704	1.0720	1.1958	1.2868	1.3653	1.4279	1.4812
SSC	0.4139	0.6775	0.8919	1.0743	1.1944	1.2825	1.3663	1.4281	1.4785
JIVE	0.2800	0.5128	0.7601	0.9038	1.0791	1.1443	1.2342	1.2866	1.3217
Single									0.2173
Pooled (10 studies)									1.0920
	experiment parameters (C=2, rho=0.5)								
SV	1.0205	1.3002	1.4406	1.5412	1.6061	1.6583	1.6960	1.7296	1.7550
SSC	0.8965	1.2572	1.4261	1.5249	1.5945	1.6503	1.6889	1.7215	1.7473
JIVE	0.7766	1.1672	1.3157	1.4180	1.4710	1.5271	1.5572	1.5975	1.6199
Single									0.5781
Pooled (10 studies)									1.5319
	experiment parameters (C=10, rho=0.5)								
SV	1.7591	1.8362	1.8794	1.9009	1.9165	1.9290	1.9372	1.9437	1.9493
SSC	1.7336	1.8221	1.8699	1.8925	1.9109	1.9243	1.9335	1.9404	1.9463
JIVE	1.5276	1.7286	1.7685	1.7935	1.8039	1.8187	1.8279	1.8357	1.8416
Single									1.5467
Pooled (10 studies)									1.8984
B. fully dependence ($\theta_{ij} = \rho\ (= 1)$ for $1 \leq i, j \leq 200$)									
	experiment parameters (C=1, rho=1)								
SV	0.4586	0.7211	0.9189	1.0850	1.2129	1.3116	1.3818	1.4428	1.4950
SSC	0.4080	0.7001	0.9011	1.0735	1.1948	1.2959	1.3748	1.4350	1.4874
JIVE	0.2806	0.5157	0.7857	0.9121	1.0551	1.1398	1.2403	1.2851	1.3265
Single									0.2403
Pooled (10 studies)									1.1129
	experiment parameters (C=2, rho=1)								
SV	1.0521	1.3144	1.4517	1.5476	1.6087	1.6554	1.6923	1.7217	1.7473
SSC	0.9641	1.2724	1.4225	1.5225	1.5915	1.6440	1.6852	1.7158	1.7428
JIVE	0.8026	1.1573	1.3321	1.4238	1.4738	1.5370	1.5609	1.6085	1.6247
Single									0.5892
Pooled (10 studies)									1.5352
	experiment parameters (C=10, rho=1)								
SV	1.7375	1.8240	1.8708	1.8958	1.9123	1.9251	1.9336	1.9404	1.9455
SSC	1.7021	1.8099	1.8615	1.8889	1.9072	1.9209	1.9302	1.9373	1.9428
JIVE	1.5482	1.7357	1.7704	1.7887	1.8087	1.8193	1.8288	1.8361	1.8434
Single									1.5188
Pooled (10 studies)									1.8912

# of studies	2	3	4	5	6	7	8	9	10
C. different sample size in the block diagonal structure (i.e., equivalent to the scenario A above)									
	experiment parameters (C=1, rho=0.5, n=40)								
SV	0.9198	1.2787	1.4549	1.5611	1.6320	1.6846	1.7198	1.7490	1.7753
SSC	0.8716	1.2177	1.4059	1.5129	1.5823	1.6348	1.6738	1.7050	1.7345
JIVE	0.3134	0.5098	0.7493	0.9808	1.0591	1.1615	1.2223	1.2921	1.3179
Single									0.4154
Pooled (10 studies)									1.5272
	experiment parameters (C=2, rho=0.5, n=40)								
SV	1.5040	1.6602	1.7435	1.7971	1.8308	1.8535	1.8701	1.8840	1.8963
SSC	1.4622	1.6357	1.7201	1.7767	1.8138	1.8385	1.8576	1.8727	1.8860
JIVE	0.7313	1.1027	1.3286	1.4158	1.4666	1.5221	1.5663	1.5967	1.6240
Single									1.0713
Pooled (10 studies)									1.7867
	experiment parameters (C=10, rho=0.5, n=40)								
SV	1.9054	1.9374	1.9528	1.9621	1.9683	1.9726	1.9759	1.9784	1.9805
SSC	1.8988	1.9346	1.9509	1.9609	1.9674	1.9718	1.9751	1.9777	1.9799
JIVE	1.4397	1.7398	1.7682	1.7900	1.8059	1.8200	1.8268	1.8360	1.8413
Single									1.8105
Pooled (10 studies)									1.9601
D. simple diagonal covariance (i.e., $\theta_{ij} = 0$).									
	experiment parameters (C=1)								
SV	0.4158	0.6702	0.8981	1.0871	1.2061	1.3098	1.3909	1.4416	1.4945
SSC	0.4203	0.6679	0.9070	1.0824	1.2009	1.2911	1.3719	1.4257	1.4768
JIVE	0.3180	0.5121	0.7591	0.9770	1.0653	1.1480	1.2292	1.2971	1.3282
Single									0.2300
Pooled (10 studies)									1.1025
	experiment parameters (C=2)								
SV	1.0182	1.2903	1.4459	1.5384	1.6078	1.6590	1.6968	1.7269	1.7514
SSC	0.9002	1.2549	1.4178	1.5201	1.5926	1.6467	1.6850	1.7162	1.7409
JIVE	0.7575	1.1233	1.3304	1.4137	1.4798	1.5272	1.5686	1.5962	1.6323
Single									0.5780
Pooled (10 studies)									1.5309
	experiment parameters (C=10)								
SV	1.7591	1.8362	1.8794	1.9009	1.9165	1.9290	1.9372	1.9437	1.9493
SSC	1.7336	1.8221	1.8699	1.8925	1.9109	1.9243	1.9335	1.9404	1.9463
JIVE	1.5276	1.7286	1.7685	1.7935	1.8039	1.8187	1.8279	1.8357	1.8416
Single									1.5467
Pooled (10 studies)									1.8984

# of studies	2	3	4	5	6	7	8	9	10
E. Different measurement scales with sample size n=20									
	experiment parameters (C=1, rho=0.5)								
SV	0.4656	0.7336	0.9589	1.1291	1.2338	1.3264	1.3981	1.4558	1.5053
SSC	0.4091	0.6961	0.9498	1.1077	1.2197	1.3067	1.3851	1.4416	1.4909
JIVE	0.3181	0.4893	0.8289	0.9691	1.0842	1.1939	1.2496	1.3033	1.3330
Single									0.2441
Pooled (10 studies)									1.1342
	experiment parameters (C=2, rho=0.5)								
SV	1.0126	1.2845	1.4414	1.5383	1.6085	1.6594	1.6999	1.7313	1.7557
SSC	0.9007	1.2457	1.4205	1.5268	1.5967	1.6483	1.6908	1.7227	1.7485
JIVE	0.8255	1.1586	1.3451	1.4311	1.5003	1.5514	1.5772	1.6093	1.6351
Single									0.5878
Pooled (10 studies)									1.5297
	experiment parameters (C=10, rho=0.5)								
SV	1.7629	1.8402	1.8790	1.9034	1.9178	1.9305	1.9393	1.9458	1.9512
SSC	1.7398	1.8305	1.8711	1.8976	1.9135	1.9264	1.9358	1.9424	1.9481
JIVE	1.5425	1.6892	1.7744	1.7932	1.8141	1.8245	1.8341	1.8425	1.8459
Single									1.5479
Pooled (10 studies)									1.9002
F. Different measurement scales with sample size n=40									
	experiment parameters (C=1, rho=0.5)								
SV	0.8255	1.1593	1.3596	1.4698	1.5514	1.6145	1.6622	1.6988	1.7281
SSC	0.7790	1.1323	1.3201	1.4246	1.5080	1.5717	1.6236	1.6603	1.6935
JIVE	0.5798	0.9749	1.2673	1.3818	1.4555	1.5092	1.5522	1.5912	1.6080
Single									0.3632
Pooled (10 studies)									1.4520
	experiment parameters (C=2, rho=0.5)								
SV	1.4095	1.5938	1.6934	1.7491	1.7898	1.8194	1.8420	1.8583	1.8725
SSC	1.3580	1.5701	1.6789	1.7356	1.7760	1.8076	1.8304	1.8471	1.8621
JIVE	1.2617	1.5070	1.5998	1.6397	1.6838	1.7115	1.7323	1.7524	1.7623
Single									0.9432
Pooled (10 studies)									1.7397
	experiment parameters (C=10, rho=0.5)								
SV	1.8808	1.9198	1.9392	1.9519	1.9604	1.9660	1.9701	1.9733	1.9759
SSC	1.8695	1.9152	1.9359	1.9497	1.9584	1.9642	1.9687	1.9722	1.9749
JIVE	1.7703	1.8207	1.8412	1.8519	1.8612	1.8662	1.8710	1.8736	1.8760
Single									1.7639
Pooled (10 studies)									1.9497

Table S8: Differential expression genes detected by muliti-class linear models. The cut-off criteria is adjusted p-value (< 0.05) by Benjamini and Hochberg's method. An R package (limma) was used to fit the models.

Data set	Brown	Heart	Liver	Skeleton
The number of DE genes	154	574	110	282

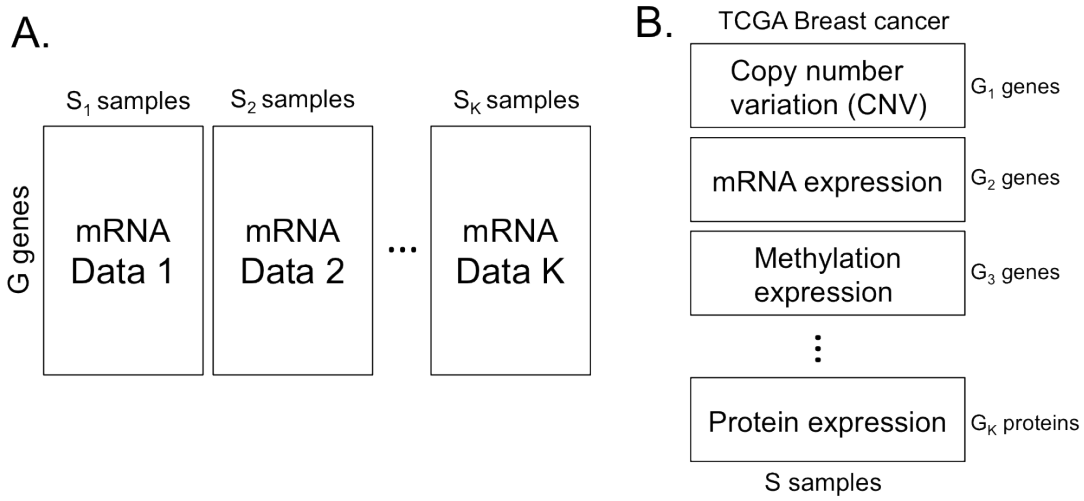

Figure S1: Two major types of omics data integration: (A) Horizontal omics meta-analysis to combine K transcriptomic datasets (B) Vertical omics integrative analysis to combine different omics data in a given cohort.

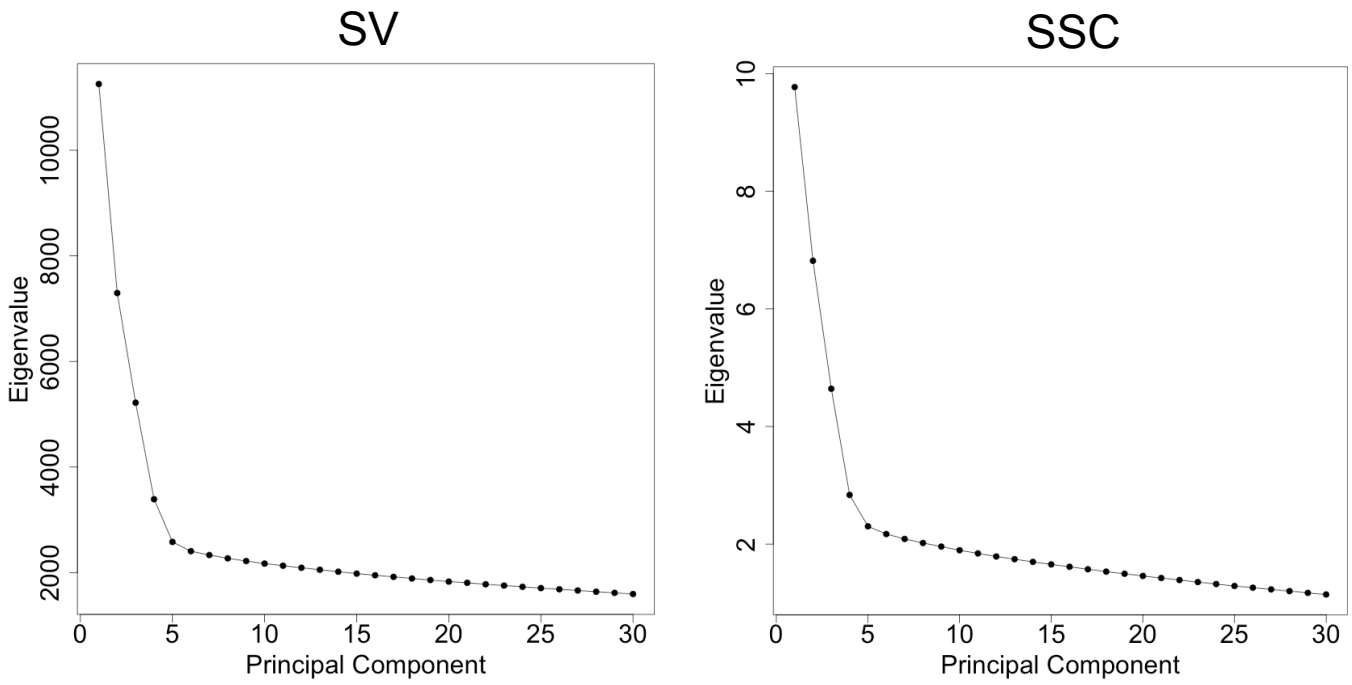

Figure S2: The example of scree plot to determine the best dimension of Meta-PCA. In scree plot, eigenvalues are sorted in decreasing order on the y-axis. Selection of the optimal K is determined by an elbow point at which the decreasing trend becomes flat when $d(i, i+1) < \Delta$, where $d(i, i+1) = \frac{e(i)-e(i+1)}{e(i)}$, $\Delta = 0.1$ and $e(i)$ refers to an eigenvalue of i^{th} leading principal component. Here we randomly generated data exploiting the similar simulation scenario in the section 3. The following is the description of the simulation scenario. Let $E = (e_1, e_2, e_3, e_4, e_5) \in \mathrm{R}^{200\times 5}$ be the true eigenvector matrix, where $e_1 = (1, 0, \ldots, 0)$, $e_2 = (0, 1, 0, \ldots, 0)$, $e_3 = (0, 0, 1, 0, \ldots, 0)$, $e_4 = (0, 0, 0, 1, 0, \ldots, 0)$, $e_5 = (0, 0, 0, 0, 1, 0, \ldots, 0)$. Denote true eigenvalues by $\lambda_1 = 500, \lambda_2 = 300, \lambda_3 = 200, \lambda_4 = 100,$ and $\lambda_5 = 50$, and create a diagonal matrix $\lambda = diag(\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5)$. Define $C = 5$ and $j^{(m)} = 5$ $(1 \leq m \leq M = 10)$. The simulations are repeated 100 times and average values are presented. For both SV and SSC in Figure S2, the elbow points declared by $d(i, i+1)$ clearly indicate the fifth principal component, and so we choose $K = 5$ as the best.

Table S9: Evaluation measures of simulated data sets and standard errors shown in Figure 3C of the main manuscript.

# of studies	2	3	4	5	6	7	8	9	10
SV	1.7591	1.8362	1.8794	1.9009	1.9165	1.9290	1.9372	1.9437	1.9493
s.e(SV)	0.0077	0.0044	0.0023	0.0019	0.0016	0.0014	0.0011	0.0008	0.0007
SSC	1.7336	1.8221	1.8699	1.8925	1.9109	1.9243	1.9335	1.9404	1.9463
s.e(SSC)	0.0086	0.0051	0.0027	0.0023	0.0019	0.0015	0.0012	0.0010	0.0009
JIVE	1.5276	1.7286	1.7685	1.7935	1.8039	1.8187	1.8279	1.8357	1.8416
s.e(JIVE)	0.0425	0.0048	0.0024	0.0021	0.0016	0.0013	0.0011	0.0011	0.0010
Single									1.5467
s.e(Single)									0.0158
Pooled (10 studies)									1.8984
s.e(Pooled)									0.0088

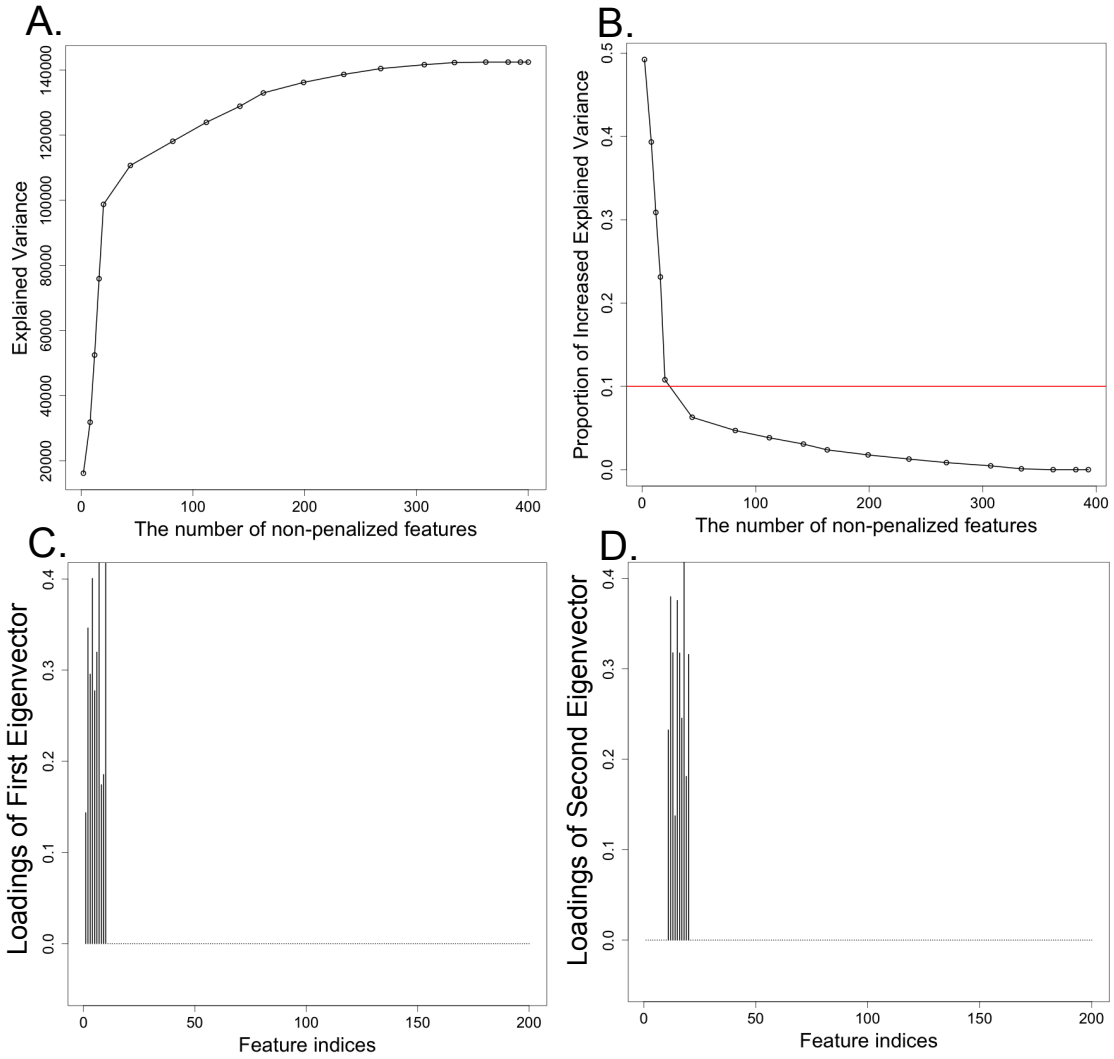

Figure S3: The example of the scree plot to determine the optimal penalization constant for MetasparsePCA. We generate simulated data sets using the same scenario as introduced in the section 3. Given fixed K, we select the best λ via the scree plot based on $G(a, b)$, a proportion of increased explained variance as a benchmark, where $G(a, b) = \frac{f(b)-f(a)}{f(b)}$ and $f(z)$ is explained variance of PC when the z number of non-zero features of eigenvector matrix are applied. Here, note that the number of non-zero features corresponds to size of λ. Two arbitrary λ values are taken to produce a and b non-zero features of eigenvector matrix such that $G(a, b) < \Delta$, where $\Delta = 0.1$, $0 \leq a, b \leq p^*$, $a < b$ and p^* the number of entire non-zero features. The simulations are repeated 100 times and average values are presented. Figure S3A-D shows that the stopping rule automatically chooses 20 nonzero features of true eigenvector matrix, suggesting that the selected penalization constant λ correctly leaves the true 20 non-zero features.

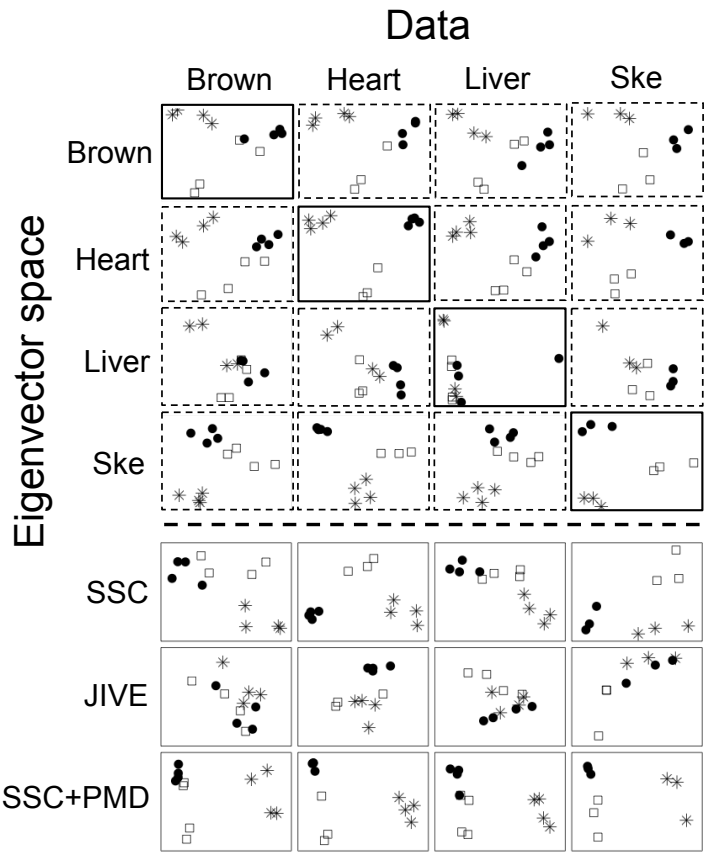

Figure S4: Two dimensional PC projections using mRNA expressions of four mouse metaboloism data; WT (square), LCAD (dot) and VLCAD (star). Brown denotes Brown fat and Ske denotes skeletal.

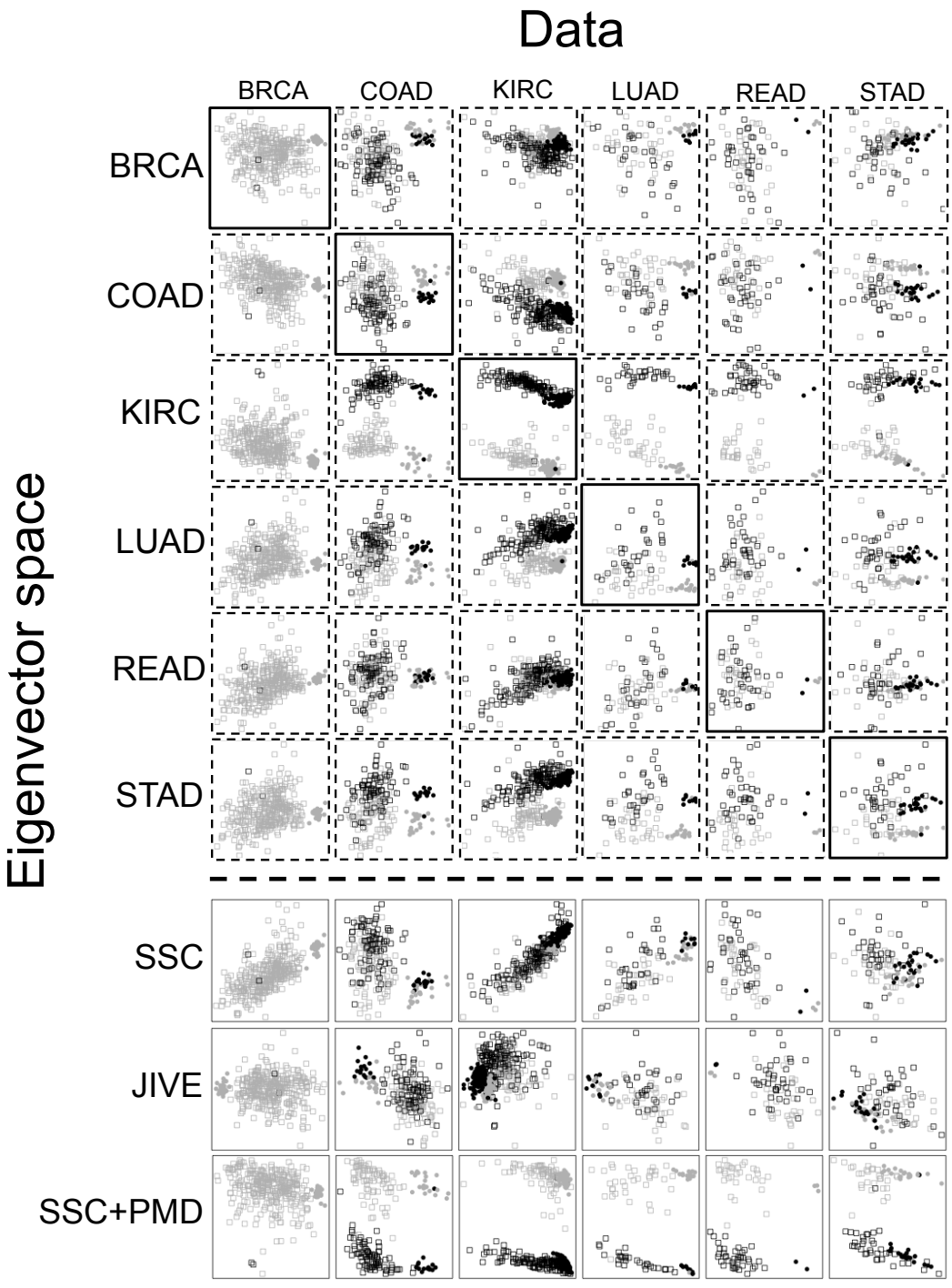

Figure S5: Two dimensional PC projections using methylation expressions of six different cancers (TCGA) data; Tumor (square), Normal (dot), Male (black) and Female (grey).

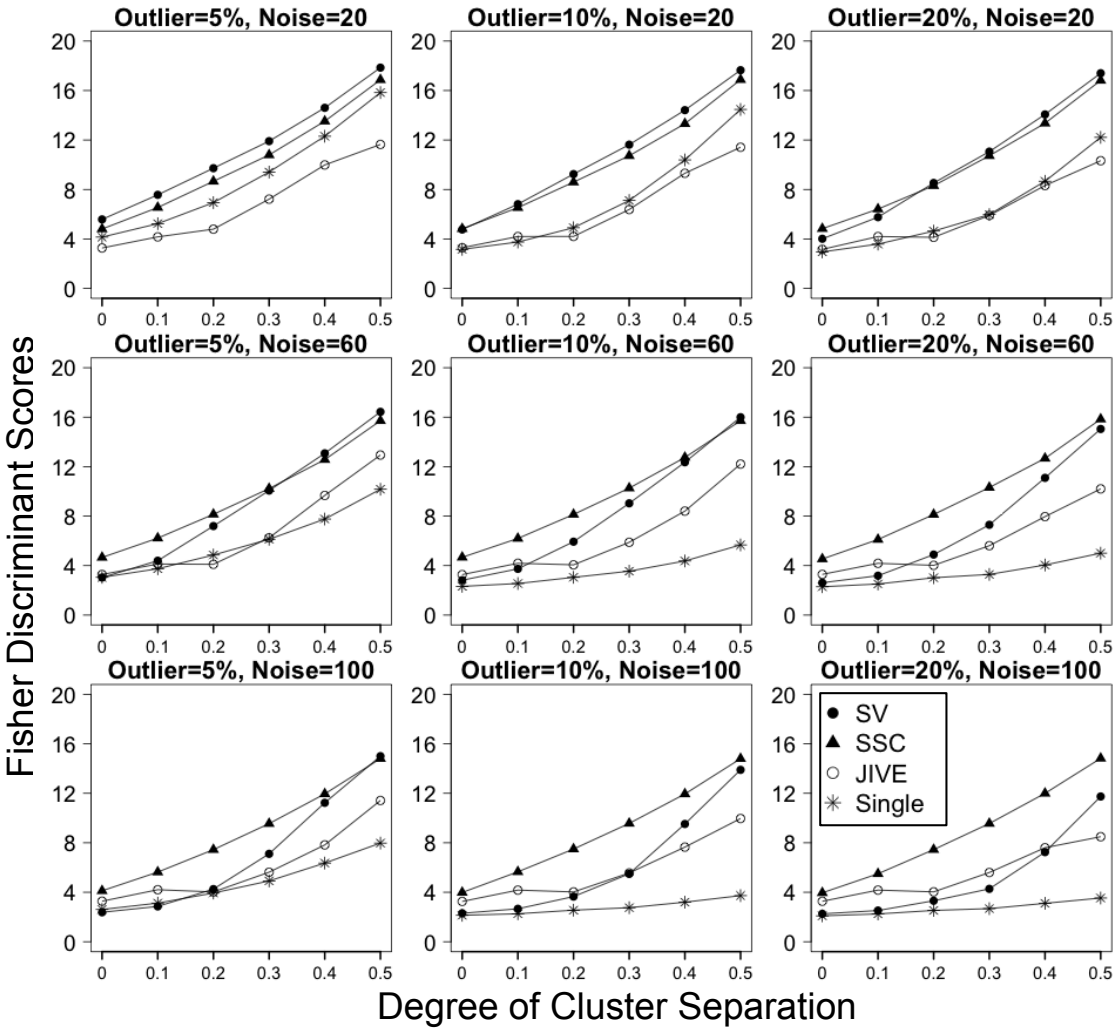

Figure S6: Robustness comparisons of MetaPCA, JIVE and PCA to outliers and noises. The y-axis represents the averages of Fisher discriminant scores, and the x-axis the magnitude of cluster separation. The figure presents the two MetaPCA methods SV (dot), SSC (triangle), JIVE (circle) and standard PCA (Single, star) applied to each individual study.

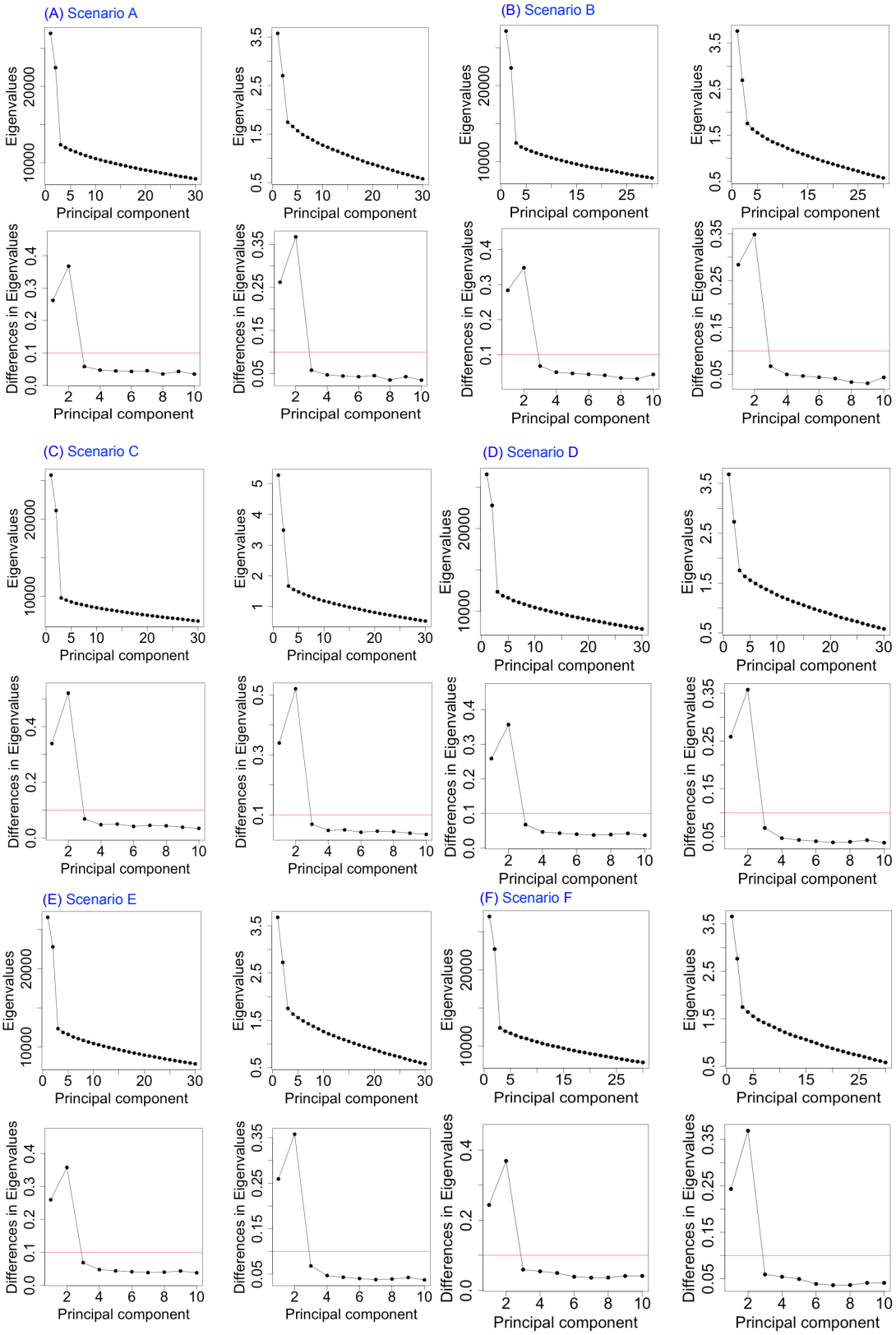

Figure S7: Scree and differences in eigenvalue plots to determine the optimal K for each scenario.

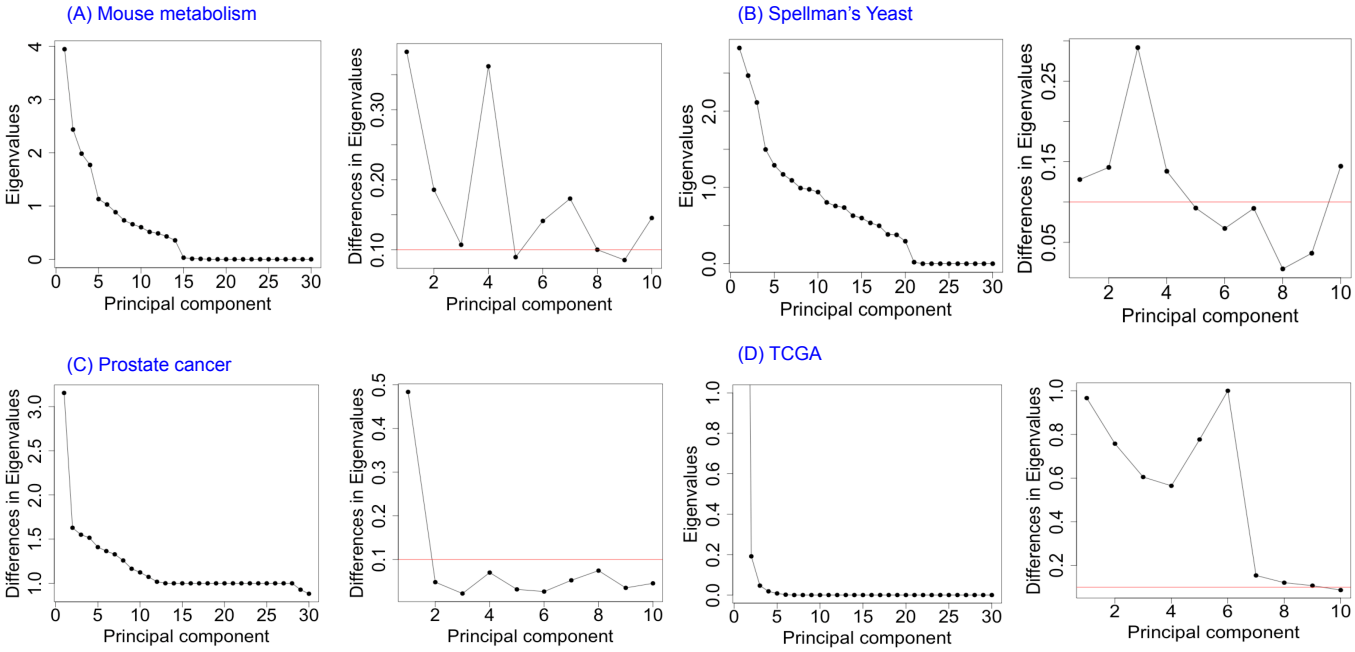

Figure S8: Scree and differences in eigenvalue plots to determine the optimal K for real data examples.