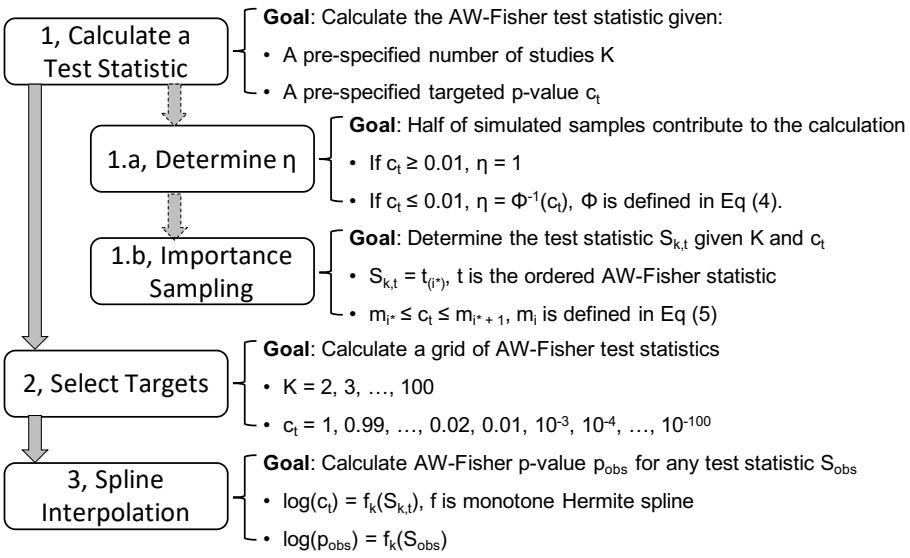


Figure S1: Summary flowchart of the AW-Fisher p-value calculation procedure based on importance sampling and spline interpolation.

Table S1: AW-Fisher p-value accuracy in terms of root mean squared error (rMSE) comparing interpolation approach, permutation-based approach, and Monte Carlo approach with closed form solution as benchmark. Two studies (sample size $N_1 = 20$, $N_2 = 20$) are included as input. B is number of permutations/samplings, and the closed form solution and the interpolation approach don't require any permutation. The range of the resulting AW-Fisher p-values are displayed in the first column. The computing time for each method is displayed in the last row. The computing time for the closed form solution is 0.03 seconds.

p-value range	Interpolation	Permutation		Monte Carlo	
		$B = 10^3$	$B = 10^4$	$B = 10^3$	$B = 10^4$
(0.01,1]	0.0002	0.0027	0.0008	0.0014	0.0002
(0.001,0.01]	0.0003	0.06	0.02	0.042	0.0047
(0.0001,0.001]	0.0007	0.32	0.1	0.29	0.045
(1e-10,0.0001]	0.0006	3.4	2.6	3.4	2.6
(1e-50,1e-10]	0.0026	20.5	19.7	20.5	19.7
(1e-100,1e-50]	0.0063	61.2	60.2	61.2	60.2
(0,1e-100]	0.72	118.0	117.0	118.0	117.0
time	0.0036 secs	11.6 mins	2.0 hours	8.2 mins	1.4 hours

Table S2: AW-Fisher p-value accuracy in terms of root mean squared error (rMSE) comparing interpolation approach, permutation-based approach, and Monte Carlo approach with closed form solution as benchmark. Two studies (sample size $N_1 = 20$, $N_2 = 20$) are included as input. B is number of permutations/samplings, and the closed form solution and the interpolation approach don't require any permutation. The range of the resulting AW-Fisher p-values are displayed in the first column. The computing time for each method is displayed in the last row. The computing time for the closed form solution is 0.03 seconds.

p-value range	Interpolation	Permutation		Monte Carlo	
		$B = 10^3$	$B = 10^4$	$B = 10^3$	$B = 10^4$
(0.01,1]	0.0002	0.0027	0.0003	0.0014	0.0002
(0.001,0.01]	0.0003	0.063	0.013	0.051	0.0058
(0.0001,0.001]	0.0007	0.32	0.072	0.31	0.049
(1e-10,0.0001]	0.0006	3.5	2.7	3.5	2.7
(1e-50,1e-10]	0.0027	22.4	21.5	22.4	21.5
(1e-100,1e-50]	0.0066	63.8	62.9	63.8	62.9
(0,1e-100]	0.9	126.3	125.3	126.3	125.3
time	0.0042 secs	12.4 mins	2.0 hours	9.8 mins	1.7 hours

Table S3: Type I error rate performance evaluation of the interpolation method via simulation studies. We simulated $K = 2, 5, 10$ studies with sample sizes (1) $N = 20$ in all studies; (2) $N = 50$ in all studies; and (3) sample sizes alternate between $N_1 = 20$ and $N_2 = 50$ in each study. 10^8 null genes (with zero effect size) were simulated in each study. The nominal α levels were pre-specified at 5.00×10^{-2}, 1.00×10^{-2}, 1.00×10^{-3}, 1.00×10^{-4}, and 2.50×10^{-6}.

(1) Sample size $N = 20$ in each study

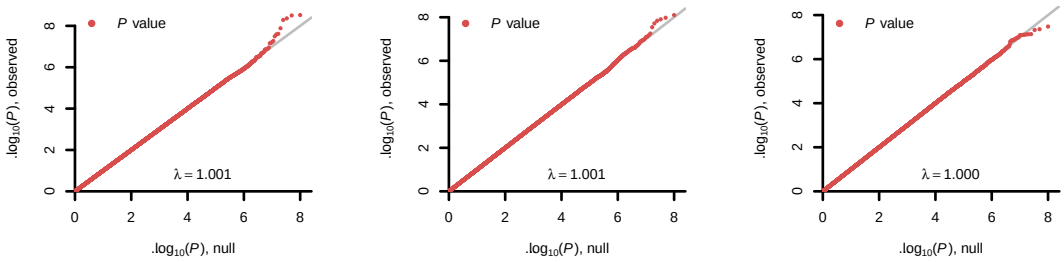
nominal α level	actual α level		
	$K = 2$	$K = 5$	$K = 10$
5.00×10^{-2}	5.01×10^{-2}	4.99×10^{-2}	4.98×10^{-2}
1.00×10^{-2}	9.98×10^{-3}	9.97×10^{-3}	9.96×10^{-3}
1.00×10^{-3}	1.00×10^{-3}	9.96×10^{-4}	9.91×10^{-4}
1.00×10^{-4}	1.00×10^{-4}	9.71×10^{-5}	1.01×10^{-4}
2.50×10^{-6}	2.38×10^{-6}	2.14×10^{-6}	2.37×10^{-6}

(2) Sample size $N = 50$ in each study

nominal α level	actual α level		
	$K = 2$	$K = 5$	$K = 10$
5.00×10^{-2}	5.00×10^{-2}	4.99×10^{-2}	4.98×10^{-2}
1.00×10^{-2}	9.96×10^{-3}	9.97×10^{-3}	9.96×10^{-3}
1.00×10^{-3}	9.93×10^{-4}	9.93×10^{-4}	9.91×10^{-4}
1.00×10^{-4}	9.85×10^{-5}	9.90×10^{-5}	9.88×10^{-5}
2.50×10^{-6}	2.48×10^{-6}	2.37×10^{-6}	2.46×10^{-6}

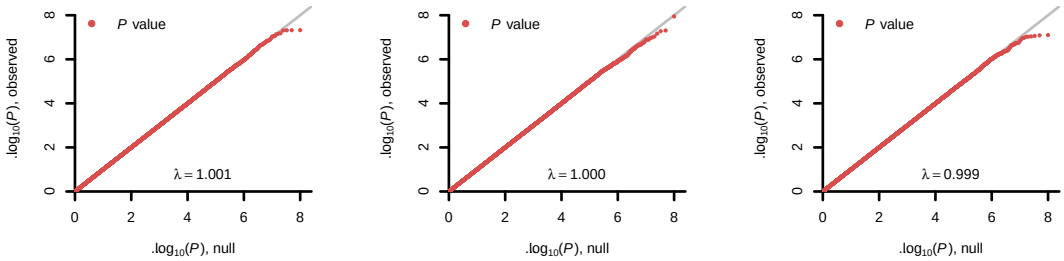
(3) Sample sizes alternate between $N_1 = 20$ and $N_2 = 50$ in each study

nominal α level	actual α level		
	$K = 2$	$K = 5$	$K = 10$
5.00×10^{-2}	5.01×10^{-2}	4.99×10^{-2}	4.98×10^{-2}
1.00×10^{-2}	9.98×10^{-3}	9.99×10^{-3}	9.97×10^{-3}
1.00×10^{-3}	1.00×10^{-3}	9.97×10^{-4}	9.91×10^{-4}
1.00×10^{-4}	1.00×10^{-4}	1.00×10^{-4}	9.98×10^{-5}
2.50×10^{-6}	2.45×10^{-6}	2.37×10^{-6}	2.49×10^{-6}



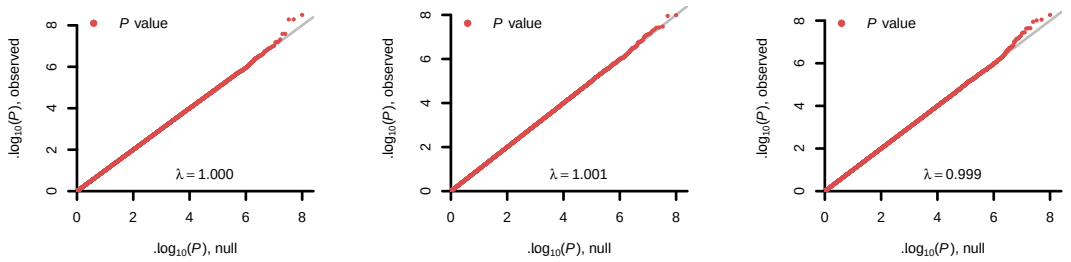
(a) K=2 (b) K=5 (c) K=10

Figure S2: QQplot of the interpolation method via simulation studies. We simulated $K = 2, 5, 10$ studies with sample sizes $N = 20$ in all studies. 10^8 null genes were simulated in each study. The genomic inflation factor λ is marked on the plot.



(a) K=2 (b) K=5 (c) K=10

Figure S3: QQplot of the interpolation method via simulation studies. We simulated $K = 2, 5, 10$ studies with sample sizes $N = 50$ in all studies. 10^8 null genes were simulated in each study. The genomic inflation factor λ is marked on the plot.



(a) K=2 (b) K=5 (c) K=10

Figure S4: QQplot of the interpolation method via simulation studies. We simulated $K = 2, 5, 10$ studies with sample sizes alternating between $N_1 = 20$ and $N_2 = 50$ in each study. 10^8 null genes were simulated in each study. The genomic inflation factor λ is marked on the plot.

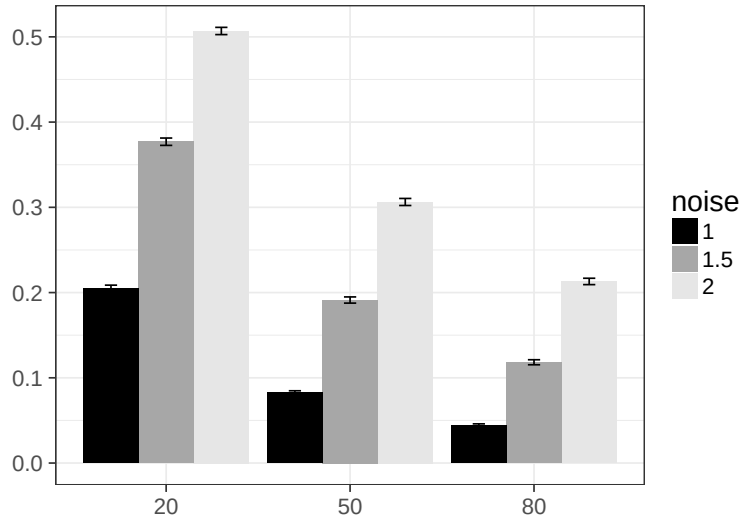


Figure S5: Comparison table of variability index for different scenarios (combinations of sample sizes ($N = 20, 50, 80$) and noise levels $\sigma = 1, 1.5, 2$). Only differential expressed genes counting from each individual studies are considered. Height of each bar indicates the mean level of variability index and error bar indicates the standard error.

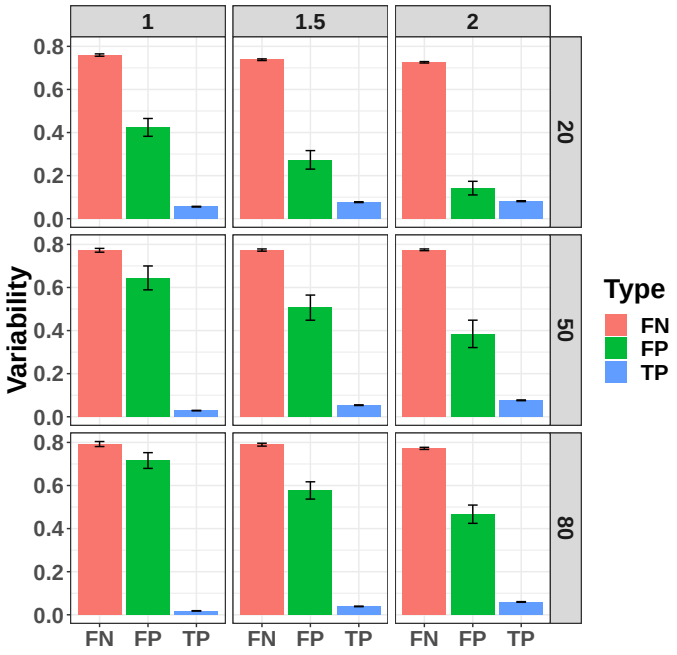


Figure S6: Variability by true positives (TP), false positives (FP), and false negatives (FN) for different combinations of sample sizes ($N = 20, 50, 80$) and noise levels $\sigma = 1, 1.5, 2$).

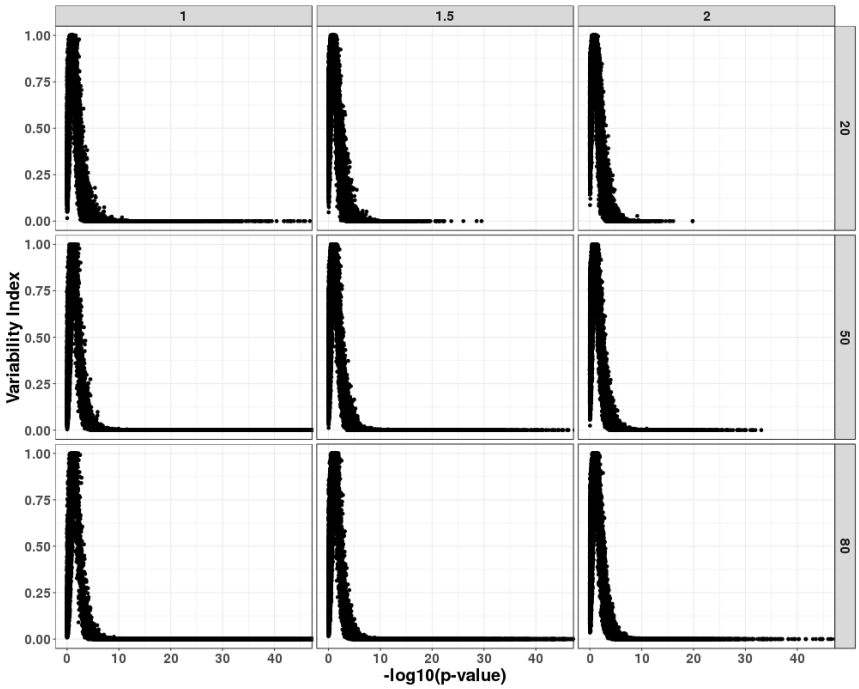


Figure S7: Scattered plot of p-value (x-axis, in $-\log_{10}$ scale) vs variability index (y-axis) for the declared DE genes (AW-Fisher q-value ≤ 0.05) for different combinations of sample sizes ($N = 20, 50, 80$) and noise levels $\sigma = 1, 1.5, 2$).

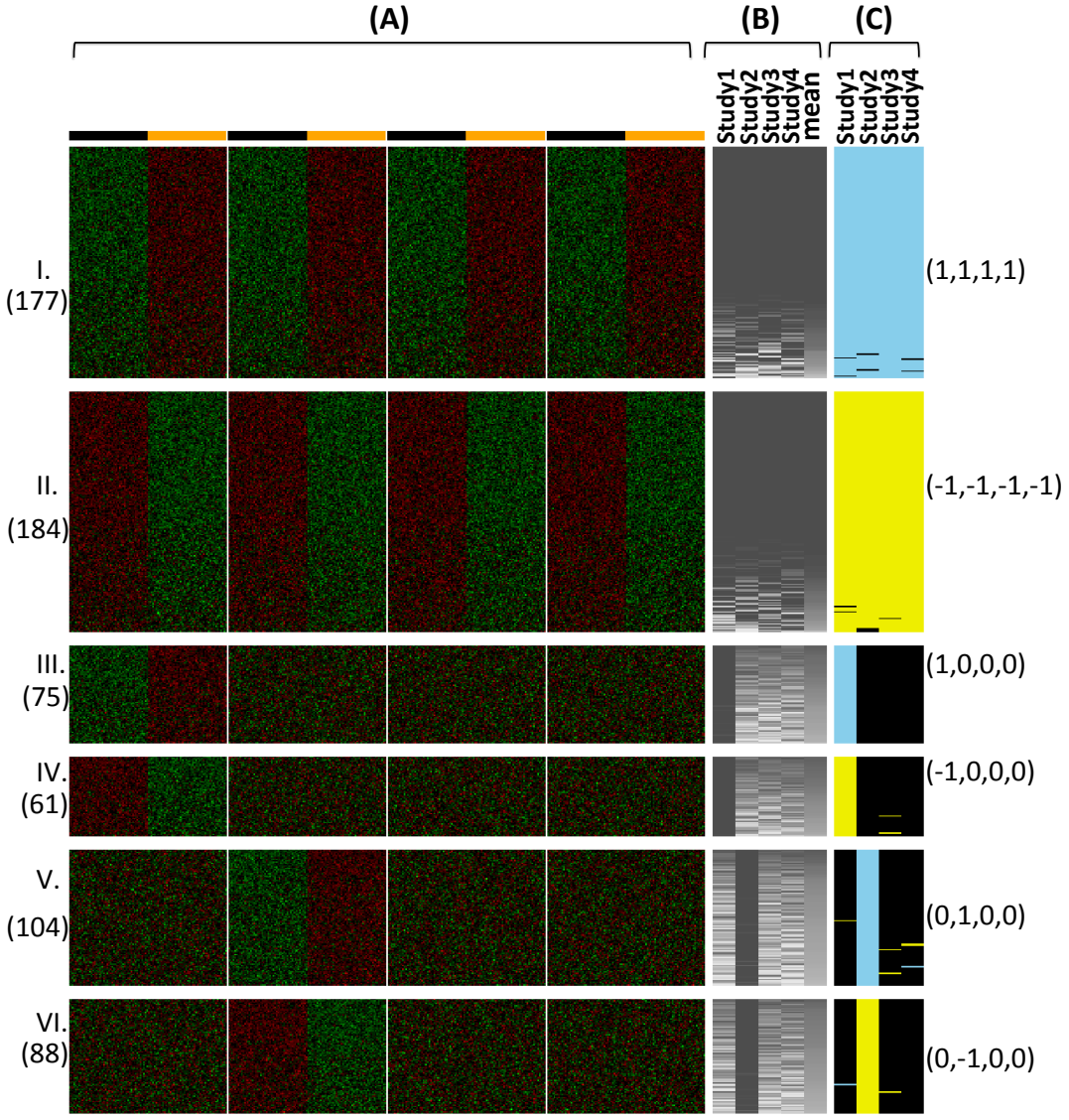


Figure S8: Six meta-pattern modules of biomarkers from the simulation study. Each gene module (Module I, II, ..., VI) shows a set of detected biomarkers with similar meta-pattern of differential signals. (A) Heatmaps of detected genes (on the rows) and samples (on the columns) for each study. In the heatmap, red color represents higher expression level, and the green color represents lower expression level. Black color bar on top represents control subjects and orange color bar on top represents case subjects. Number of genes is shown on the left under each module number. (B) Variability index (genes on the rows and studies on the columns). Variability index is described in Section 2.1 Gray heatmap range from 0 (black) to 1 (white), which is the maximum of the variability index. Genes of each module are sorted based on the mean variability index. (C) Signed AW-Fisher weights $\hat{v}_{gk}$ for gene g and study k. Light blue represents $\hat{v}_{gk} = 1$, yellow corresponds to $\hat{v}_{gk} = -1$ and black for $\hat{v}_{gk} = 0$. Representative signed AW-Fisher weights for each module are shown on the right.

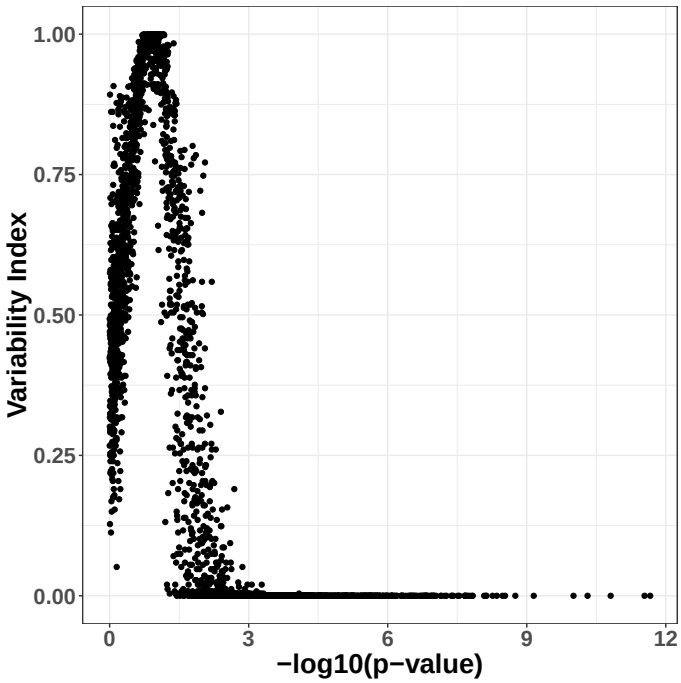


Figure S9: Scattered plot of p-value (x-axis, in $-\log_{10}$ scale) vs variability index (y-axis) for the declared DE genes (AW-Fisher q-value ≤ 0.05) of the mouse metabolism data.

Table S4: Contingency table of 794 detected DE genes with simulation underlying truth (on the columns) and the hierarchical clustering result with 7 target modules (on the rows). 0 represents the scattered gene group. $1 \sim 6$ represent 6 detected modules. Bolded numbers are genes with correct assignment.

Module	homo−	homo+	ssp1−	ssp1+	ssp2−	ssp2+	nonDE
1	0	**160**	0	0	0	0	0
2	22	41	20	1	0	2	**22**
3	**175**	0	0	0	0	0	0
4	0	0	0	0	0	**105**	4
5	0	0	0	0	**91**	0	4
6	0	0	0	**84**	0	0	4
0	0	0	**59**	0	0	0	0

Table S5: Contingency table of 794 detected DE genes with simulation underlying truth (on the columns) and K-means clustering result with 7 target modules (on the rows). $1 \sim 7$ represent 7 detected modules. Bolded numbers are genes with correct assignment.

Module	homo−	homo+	ssp1−	ssp1+	ssp2−	ssp2+	nonDE
1	0	0	0	0	**87**	0	3
2	20	35	4	5	4	6	**22**
3	0	**166**	0	0	0	0	0
4	0	0	0	0	0	**101**	2
5	0	0	**75**	0	0	0	4
6	**177**	0	0	0	0	0	0
0	0	0	0	**80**	0	0	3

Table S6: Sample size summary for two real data examples. For multi-tissue microarray studies using metabolism related knockout mice, there are three tissues to be meta-analyzed comparing wild type versus VLCAD-deficient. For multi-brain-region RNA-seq studies using HIV-1 transgenic rats, there are three brain regions to be meta-analyzed comparing F334 rats versus HIV infected rats. The number in the parentheses is the number of samples actually after removing potential outliers.

(a) sample size of mouse metabolism microarray RNA expression

Tissue	wild type	VLCAD-deficient
Brown	4	4
Heart	3	4
Liver	4	4
Skeleton	3	3

(b) sample size of RNA-seq in brains of HIV-1 transgenic rat

Brain Region	F334	HIV
HIP	12	12(11)
PFC	12	12(10)
STR	12	12

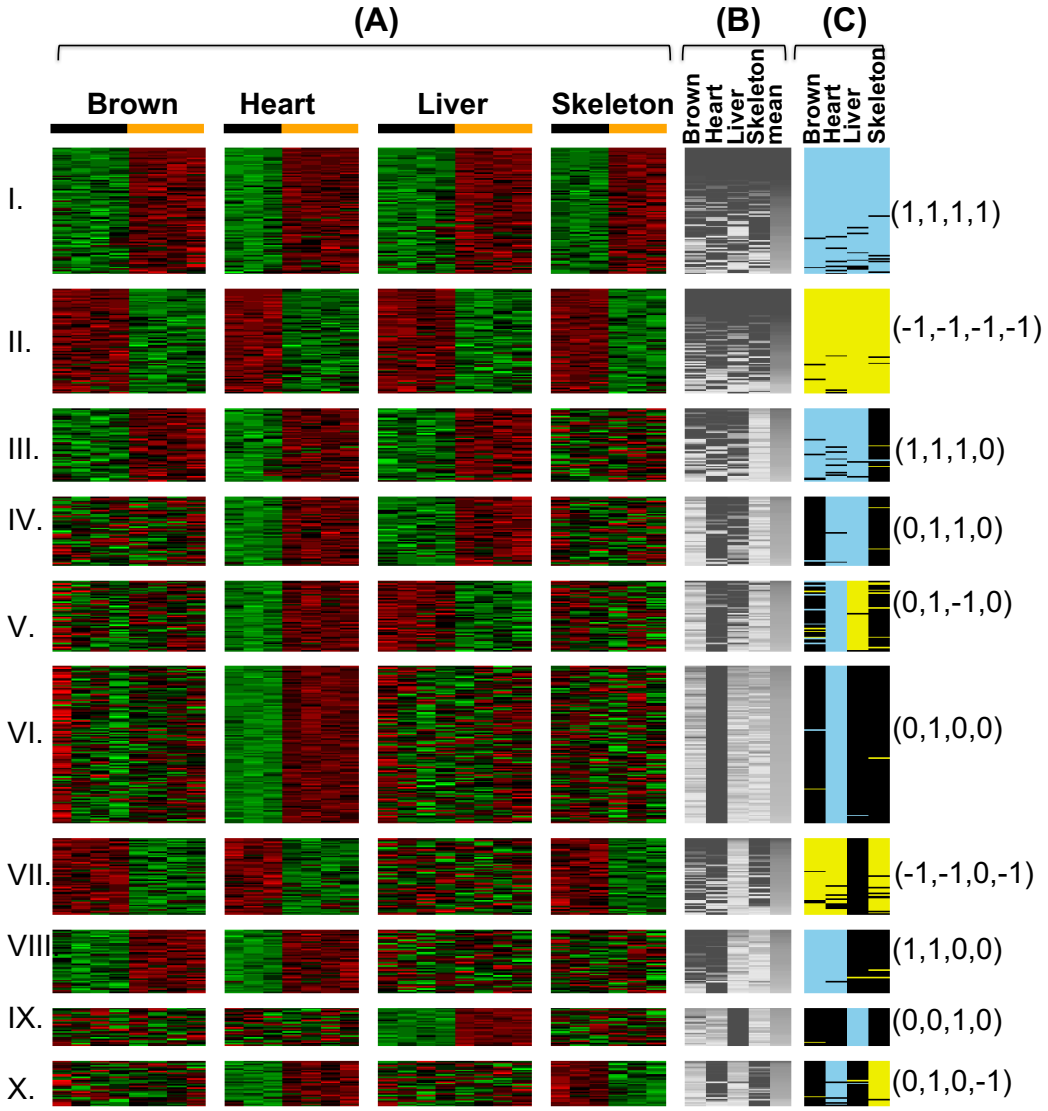
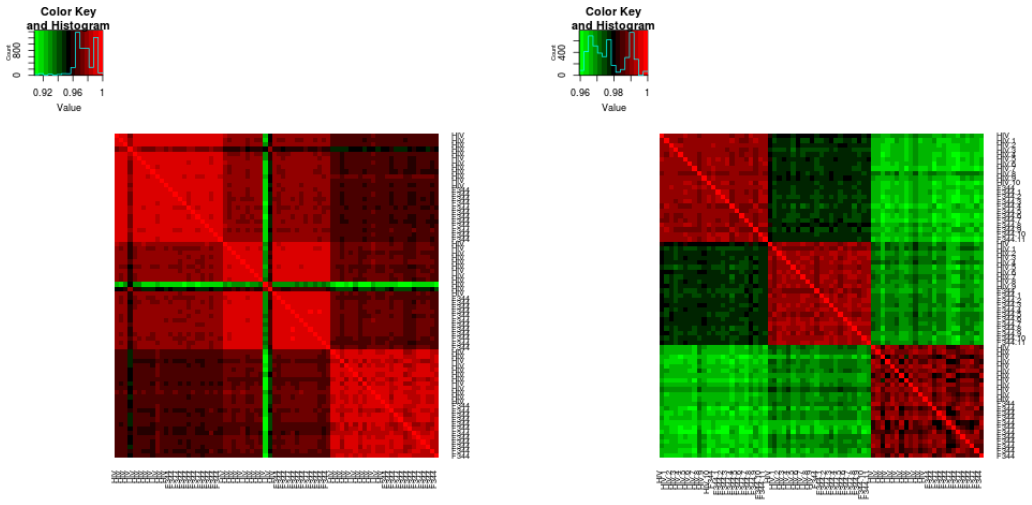


Figure S10: Ten meta-pattern modules of biomarkers from mouse metabolism example. Each gene module (Module I, II, ..., X) shows a set of detected biomarkers with similar meta-pattern of differential signals. (A) Heatmaps of detected genes (on the rows) and samples (on the columns) for each tissue (brown fat, heart, liver and skeleton), where each tissue represents a study. In the heatmap, red color represents higher expression level, and the green color represents lower expression level. Black color bar on top represents wild type (control) and orange color bar on top represents VLCAD -/- mice (case). Number of genes is shown on the left under each module number. (B) Variability index (genes on the rows and studies on the columns). Gray heatmap range from 0 (black) to 1 (white), which is the maximum of the variability index. Genes of each module are sorted based on the mean variability index. (C) Signed AW-Fisher weights $\hat{v}_{gk}$ for gene g and study k. Light blue represents $\hat{v}_{gk} = 1$, yellow corresponds to $\hat{v}_{gk} = -1$ and black for $\hat{v}_{gk} = 0$. Representative signed AW-Fisher weights for each module are shown on the right. Note Brown represents brown fat tissue.



(a) Before (b) After

Figure S11: Correlation heatmap before and after removing the outliers. Three blocks are three brain region: hippocampus (HIP), prefrontal cortex (PFC) and striatum (STR) respectively.

Table S7: Unstable AW weight estimates and their variability indexes. The first row represents the Affymetrix probe names. The p-values from the three tissues (brown fat, heart and liver) are listed in column 2 – 4. The AW weight estimates for the three tissues are listed in column 5 – 7. and the variability of AW weight are listed in column 8 – 10.

s

	p-value			weight			variability		
Probe	Brown	Heart	Liver	Brown	Heart	Liver	Brown	Heart	Liver
1419484_a_at	3.9e-04	9.6e-02	2.1e-03	1	1	1	0.00	0.93	0.00
1421163_a_at	2.3e-06	1.1e-01	8.8e-03	1	0	1	0.00	0.84	0.77
1421704_a_at	1.7e-02	7.5e-02	8.9e-05	1	1	1	0.11	0.76	0.00
1424007_at	2.7e-02	8.8e-02	2.2e-03	1	1	1	0.50	0.97	0.00
1425567_a_at	3.6e-04	1.0e-01	2.1e-03	1	0	1	0.00	0.94	0.00
1425806_a_at	1.4e-02	1.0e-01	9.9e-04	1	1	1	0.48	0.98	0.00
1429054_at	2.8e-02	1.0e-01	7.4e-04	1	1	1	0.39	0.77	0.00
1437103_at	3.8e-03	1.0e-01	1.0e-03	1	1	1	0.03	0.99	0.00
1448028_at	5.7e-05	1.1e-01	5.1e-03	1	0	1	0.00	1.00	0.26
1449518_at	1.8e-02	1.0e-01	1.2e-03	1	1	1	0.30	0.99	0.00
1452418_at	6.5e-07	1.1e-01	3.5e-03	1	0	1	0.00	0.93	0.03

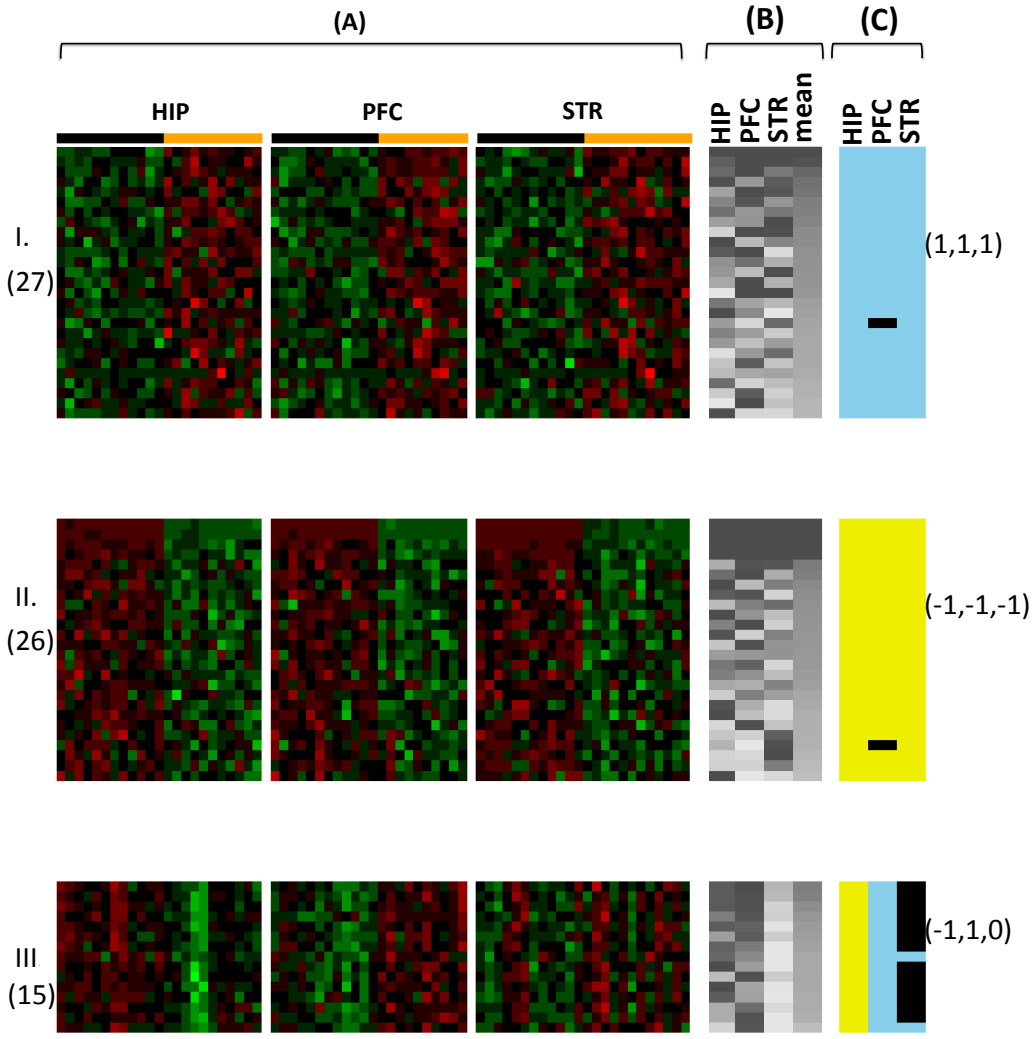


Figure S12: Three meta-pattern modules of biomarkers from HIV transgenic rats example. Each gene module (Module I, II and III) shows a set of detected biomarkers with similar meta-pattern of differential signals. (A) Heatmaps of detected genes (on the rows) and samples (on the columns) for each brain region (HIP, PFC or STR). where each brain region represents a study k. In the heatmap, red color represents higher expression level, and the green color represents lower expression level. Black color bar on top represents F334 rats (control) and orange color bar on top represents HIV transgenic rats (case). Number of genes is shown on the left under each module number. (B) Variability index (genes on the rows and studies on the columns). Variability index is described in Section 2.1, Gene modules, gray heatmap range from 0 (black) to 1 (white), which is the maximum of the variability index. Genes of each module are sorted based on the mean variability index. (C) AW weight result. Light blue color represents AW weight 1 and up-regulation. Yellow color represents AW weight 1 and down-regulation. Black color represents AW weight 0. Genes are shown on the rows and studies are shown on the columns Number of genes is shown on right of each module.

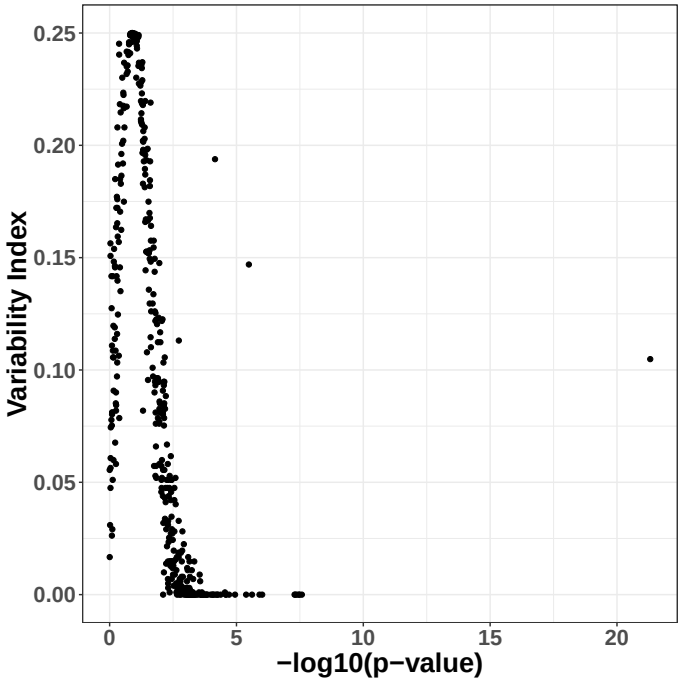


Figure S13: Scattered plot of p-value (x-axis, in $-\log_{10}$ scale) vs variability index (y-axis) for the declared DE genes (AW-Fisher q-value ≤ 0.05) of the mouse HIV data.

I closed-form solution for AW-Fisher K=2 and K=3

I.1 K=2

Recall that $S = \min_{\vec{\mathbf{w}}} L(T(\vec{\mathbf{w}}; \vec{\mathbf{P}}))$ and the observed AW-Fisher statistics is $s_{obs} = s(\vec{\mathbf{p}}_{obs})$. Let $t_j = \exp(-\chi^{-2}_{2j}(s_{obs})/2)$ and $T_j = \prod_{i=1}^{j} P_{(i)}$, where $\chi^{-2}_{2j}(t)$ is the $100(1-t)\%$ quantile of χ^2_{2j} and $P_{(i)}$'s are order p-values, then

$$\mathbb{P}(S \leq s_{obs}) = \mathbb{P}(\cup_{j=1}^{K}\{T_j < t_j\}). \tag{1}$$

Since the joint distribution of $P_{(i)},\ i = 1, \ldots, K$ is $f(p_1, \ldots, p_K) = K!,\ 0 \leq P_{(1)} \leq P_{(2)} \leq \cdots \leq P_{(K)} \leq 1$, it is possible to estimate $P(S \leq s_{obs})$ analytically. Without loss of generality, let's denote by A_j the event $\{T_j < t_j\}$, then $\mathbb{P}(S \leq s_{obs})$ can be rewritten as

$$\mathbb{P}(S \leq s_{obs}) = \mathbb{P}(\cup_{j=1}^{K} A_j) = \mathbb{P}(A_1) + \sum_{k=2}^{K} \mathbb{P}(A_k \cap (\cup_{j=1}^{k-1} A_j^c)), \tag{2}$$

where A_j^c is the complementary event of A_j. The above formula provides an analytical way to compute the p-value $\mathbb{P}(S \leq s_{obs})$. For example, when $K = 2$,

$$\begin{aligned}
\mathbb{P}(S \leq s_{obs}) &= \mathbb{P}(\cup_{j=1}^{2}\{T_j < t_j\}) = 1 - \mathbb{P}(T_1 \geq t_1, T_2 \geq t_2) \\
&= 1 - \int_{t_1}^{1}\int_{t_2/p_1}^{1} 2 \cdot \mathbb{1}(p_1 \leq p_2) dp_2 dp_1 = 1 - \int_{t_1}^{1}\int_{\max\{t_2/p_1, p_1\}}^{1} 2 dp_2 dp_1 \\
&= 1 - \int_{\max(t_1, t_2^{1/2})}^{1} 2(1-p_1)dp_1 - \int_{\min(t_1, t_2^{1/2})}^{t_2^{1/2}} 2(1 - t_2/p_1)dp_1 \\
&= 1 + (1-p_1)^2 \mid_{\max(t_1, t_2^{1/2})}^{1} -2(p_1 - t_2\log(p_1)) \mid_{\min(t_1, t_2^{1/2})}^{t_2^{1/2}}
\end{aligned}$$

In the case $t_1 \geq t_2^{1/2}$,

$$\mathbb{P}(S \leq s_{obs}) = 1 - (1-t_1)^2 = 2t_1 - t_2^2$$

If $t_1 < t_2^{1/2}$,

$$\begin{aligned}
\mathbb{P}(S \leq s_{obs}) &= 1 - (1 - t_2^{1/2})^2 - 2(t_2^{1/2} - t_1) + 2t_2\log(\frac{t_2^{1/2}}{t_1}) \\
&= t_2\log(t_2/t_1^2) + 2t_1 - t_2
\end{aligned}$$

Therefore, for $K = 2$, the p-value $\mathbb{P}(S \leq s_{obs})$ for given observed test statistic $-\log(t)$ can be computed analytically by

$$\mathbb{P}(S \leq s_{obs}) = \begin{cases} 2t_1 - t_1^2 & t_1^2 \geq t_2 \\ t_2\log(t_2/t_1^2) + 2t_1 - t_2, & t_1^2 < t_2 \end{cases}. \tag{3}$$

I.2 K=3

When $K = 3$,

$$
\begin{aligned}
&\mathbb{P}(S \le s_{obs}) \\
&= \mathbb{P}(\cup_{j=1}^{3}\{T_j < t_j\}) = 1 - \mathbb{P}(T_1 \ge t_1, T_2 \ge t_2, T_3 \ge t_3) \\
&= 1 - \int_{t_1}^{1}\int_{t_2/p_1}^{1}\int_{t_3/p_1p_2}^{1} 6 \cdot \mathbb{1}(p_1 \le p_2 \le p_3) dp_3 dp_2 dp_1 \\
&= 1 - 6\int_{t_1}^{1}\int_{\max(t_2/p_1, p_1)}^{1}\int_{\max(t_3/p_1p_2, p_1)}^{1} dp_3 dp_2 dp_1 \\
&= 1 - 6\int_{t_1}^{1}\int_{\max((t_3/p_1)^{1/2}, t_2/p_1, p_1)}^{1} (1-p_2) dp_2 dp_1 - 6\int_{t_1}^{1}\int_{\max(t_2/p_1, p_1)}^{\max((t_3/p_1)^{1/2}, t_2/p_1, p_1)} (1 - t_3/p_1p_2) dp_2 dp_1 \\
&= 1 + 3\int_{t_1}^{1}(1-p_2)^2 \mid_{p_2=\max((t_3/p_1)^{1/2}, t_2/p_1, p_1)}^{1} dp_1 - 6\int_{t_1}^{1}(p_2 - (t_3/p_1)\log p_2) \mid_{p_2=\max(t_2/p_1, p_1)}^{\max((t_3/p_1)^{1/2}, t_2/p_1, p_1)} dp_1 \\
&= 1 - 3\int_{\max(t_1, t_2^{1/2}, t_3^{1/3})}^{1}(1-p_1)^2 dp_1 - 3\int_{\max(t_1, \min(t_2^{1/2}, t_2^2/t_3))}^{\min(t_2^{1/2}, t_2^2/t_3)}(1 - t_2/p_1)^2 dp_1 \\
&- 3\int_{\min(t_3, \max(t_1, t_2^2/t_3))}^{t_3}(1 - (t_3/p_1)^{1/2})^2 dp_1 \\
&- 6\int_{\min(t_1, t_2^{1/2}, t_3^{1/3})}^{\min(t_2^{1/2}, t_3^{1/3})}(t_3/p_1)^{1/2} - t_2/p_1 - (t_3/2p_1)\log(t_3p_1/t_2^2) dp_1 \\
&- 6\int_{\min(t_3^{1/3}, \max(t_1, t_2^{1/2}))}^{t_3^{1/3}}(t_3/p_1)^{1/2} - p_1 - (t_3/2p_1)\log(t_3/p_1^3) dp_1
\end{aligned}
$$

The above formula can be further simplified based on the magnitude relationship among $(t_1, t_2^{1/2}, t_3^{1/3})$. There will be 5 conditional formula. We omit the mathematical details. In general, when we have K studies, there will be $\mathcal{O}(K!)$ conditional formula according the magnitude relationship among $(t_1, t_2^{1/2}, \ldots, t_K^{1/K})$.

II Simulation

The main simulation setting mimics the reality of a transcriptomic study by considering generative process of DE genes and correlation structures between genes. Note that this simulation also applies to Section 3.1.1 and Section 3.1.3. Below are the details for the simulation.

1. Simulate $K = 2$ studies, $G = 10,000$ genes and 2N subjects ($N = 20$) with N cases and N controls.

2. Firstly, we simulated correlated gene structure and assumed no effect size for any gene or any study.

   (a) For the first 4,000 genes, simulate 200 gene modules with 20 genes in each module and the remaining 6,000 genes are uncorrelated. Denote by

$C_g \in \{0, 1, \ldots, 200\}$ the cluster membership indicator for gene g (e.g., $C_g = 1$ indicates gene g is in module 1 while $C_g = 0$ indicates gene g is not in any gene module).

(b) For module c and study k, simulate $A'_{ck} \sim W^{-1}(\Phi, 60)$, where $1 \leq c \leq 200$, $\Phi = 0.5I_{20\times 20} + 0.5J_{20\times 20}$, W^{-1} denotes the inverse Wishart distribution, I is the identity matrix and J is the matrix with all elements equal to 1. A_{ck} is calculated by standardizing A'_{ck} such that the diagonal elements are all 1's. The covariance matrix for gene module c in study k is calculated as $\Sigma_{ck} = A_{ck}$.

(c) Denote by $g_{c1}, \ldots, g_{c20}$ the indices of the 20 genes in module c (i.e., $C_{g_{cj}} = c$, where $1 \leq c \leq 200$ and $1 \leq j \leq 20$). Simulate expression levels of genes in module c for sample n in study k as $(X'_{g_{c1}kn}, \ldots, X'_{g_{c20}kn}) \sim \text{MVN}(0, \Sigma_{cs})$, where $1 \leq n \leq 2N$ and $1 \leq k \leq K$. For any uncorrelated gene g with $C_g = 0$, simulate the expression level for sample n in study k as $X'_{gkn} \sim \text{N}(0, \sigma^2)$, where $1 \leq n \leq 2N$ and $1 \leq k \leq K$.

3. Simulate effect sizes and their DE directions for differentially expressed (DE) genes.

(a) Assume that the first G_1 genes are DE in at least one of the combined studies, where $G_1 = 30\% \times G$. For each $1 \leq g \leq G_1$, simulate v_g from discrete uniform distribution $v_g \sim \text{UNIF}(1, \ldots, K)$ and then randomly simulate subset $\mathbf{v}_g \subseteq \{1, \ldots, K\}$ such that $|\mathbf{v}_g| = v_g$. Here $\mathbf{v}_g$ is the set of studies in which gene g is DE.

(b) For any DE gene $g (1 \leq g \leq G_1)$, simulate gene-level effect size $\theta_g \sim \text{N}_{0.5+}(1, 1)$, where N_{a+} denotes the truncated Gaussian distribution within interval (a, ∞). Also simulate study-specific random effect size $\theta_{gk} \sim \text{N}(\theta_g, 0.2^2)$.

(c) Simulate $d_g \sim \text{BIN}(1, 0.5)$, where $1 \leq g \leq G_1$. Here d_g is the DE direction for gene g for majority of studies.

4. Add the directed effect sizes to the gene expression levels simulated in Step **??**. For control subjects $(1 \leq n \leq N)$, set the expression levels as $X_{gkn} = X'_{gkn}$. For case subjects $(N + 1 \leq n \leq 2N)$, if $1 \leq g \leq G_1$ and $k \in \mathbf{v}_g$, we set the expression levels as $X_{gkn} = X'_{gkn} + (-1)^{d_g}\theta_{gk}$.