

**Additional file 3: Supplementary Text for
“FSBLUP: A novel strategy of fusion similarity matrix
construction via optimally integrating intermediate
omics data to enhance genomic prediction”**

Yahui Xue¹, Lei Zhou¹, Yue Zhuo¹, Weining Li², Sijia Ma¹, Heng Du¹, Wanying Li¹,
Jicai Jiang³, Jian-Feng Liu^{1*}

¹ State Key Laboratory of Animal Biotech Breeding, Frontiers Science Center for Molecular Design Breeding (MOE), College of Animal Science and Technology, China Agricultural University, Beijing 100193, China

² Institute of Animal Science, Jiangsu Academy of Agricultural Sciences, Nanjing, Jiangsu 210014, China

³ Department of Animal Science, North Carolina State University, Raleigh, NC 27695, United States

* Corresponding E-mail: liujf@cau.edu.cn

Contacts:

Yahui Xue: xue_yh@cau.edu.cn

Lei Zhou: leiz@cau.edu.cn

Yue Zhuo: zhuoy@cau.edu.cn

Weining Li: liwn@jaas.ac.cn

Sijia Ma: masijia626@cau.edu.cn

Heng Du: kimi-du@cau.edu.cn

Wanying Li: liwanyingcau@foxmail.com

Jicai Jiang: jjiang26@ncsu.edu

Jian-Feng Liu: liujf@cau.edu.cn

Expanded algorithmic derivation details

Normally, $\mathbf{A}$ was used to represent the expected genetic relationships and to $\mathbf{G}$ as realized relationships; however, $\mathbf{G}$ is observed only for some individuals, and those individuals have $\mathbf{A}_{22}$ as a priori relationship. Based on that, it was shown that the genomic information could be extended to non-genotyped animal based on the joint distribution of breeding values of non-genotyped ($\mathbf{u}_1$) and genotyped ($\mathbf{u}_2$) individuals:

$$p(\mathbf{u}_1, \mathbf{u}_2) = p(\mathbf{u}_2)p(\mathbf{u}_1|\mathbf{u}_2) \quad (1)$$

$$p(\mathbf{u}_2) = N(\mathbf{0}, \mathbf{G}) \quad (2)$$

$$Var(\mathbf{u}) = \mathbf{A}\sigma_u^2 \quad (3)$$

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \quad (4)$$

where subscripts 1 and 2 represent non-genotyped and genotyped individuals, respectively. Since the intermediate omics data was incorporated, that changes the classification numbers into four types:

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} & \mathbf{A}_{13} & \mathbf{A}_{14} \\ \mathbf{A}_{21} & \mathbf{A}_{22} & \mathbf{A}_{23} & \mathbf{A}_{24} \\ \mathbf{A}_{31} & \mathbf{A}_{32} & \mathbf{A}_{33} & \mathbf{A}_{34} \\ \mathbf{A}_{41} & \mathbf{A}_{42} & \mathbf{A}_{43} & \mathbf{A}_{44} \end{bmatrix} \quad (5)$$

We extended genomic and intermediate omics information to non-measured individuals using joint distributions of breeding values to address this. Group 1 included individuals with only pedigree information, while Group 4 comprised individuals with all data types. Groups 2 and 3 included individuals lacking omics or

genomic information, respectively. For example, the joint distribution of the breeding values of $\mathbf{u}_1$ and $\mathbf{u}_4$ is:

$$p(\mathbf{u}_1, \mathbf{u}_4) = p(\mathbf{u}_4)p(\mathbf{u}_1|\mathbf{u}_4) \quad (6)$$

$$p(\mathbf{u}_4) = N(\mathbf{0}, \mathbf{C}) \quad (7)$$

The conditional distribution of breeding values for Group 1 and Group 4 is:

$$p(\mathbf{u}_1|\mathbf{u}_4) = N(\mathbf{A}_{14}\mathbf{A}_{44}^{-1}\mathbf{u}_4, \mathbf{A}_{11} - \mathbf{A}_{14}\mathbf{A}_{44}^{-1}\mathbf{A}_{41}) \quad (8)$$

It implies that by using $\mathbf{A}_{14}\mathbf{A}_{44}^{-1}\mathbf{u}_4$, the genomic and intermediate omics information can be implicitly imputed from individuals in Group 4 to Group 1 based on pedigree relationships. Besides, $\mathbf{u}_1$ could be expressed as below (see *Likelihood, Bayesian and MCMC Methods in Quantitative Genetics* [66] p254):

$$\begin{aligned} \mathbf{u}_1 &= E(\mathbf{u}_1|\mathbf{u}_4) + \boldsymbol{\varepsilon} \\ &= E(\mathbf{u}_1) + \mathbf{A}_{14}\mathbf{A}_{44}^{-1}[\mathbf{u}_4 - E(\mathbf{u}_4)] + \boldsymbol{\varepsilon} \\ &= \mathbf{A}_{14}\mathbf{A}_{44}^{-1}\mathbf{u}_4 + \boldsymbol{\varepsilon} \end{aligned} \quad (9)$$

$$\begin{aligned} \text{Var}(\boldsymbol{\varepsilon}) &= \text{Var}(\mathbf{u}_1 - \mathbf{A}_{14}\mathbf{A}_{44}^{-1}\mathbf{u}_4) \\ &= \text{Var}(\mathbf{u}_1) - \text{Var}(\mathbf{A}_{14}\mathbf{A}_{44}^{-1}\mathbf{u}_4) \\ &= \mathbf{A}_{11} - E\left[(\mathbf{A}_{14}\mathbf{A}_{44}^{-1}\mathbf{u}_4 - E[\mathbf{A}_{14}\mathbf{A}_{44}^{-1}\mathbf{u}_4])(\mathbf{A}_{14}\mathbf{A}_{44}^{-1}\mathbf{u}_4 - E[\mathbf{A}_{14}\mathbf{A}_{44}^{-1}\mathbf{u}_4])^T\right] \\ &= \mathbf{A}_{11} - E\left[\mathbf{A}_{14}\mathbf{A}_{44}^{-1}(\mathbf{u}_4 - E(\mathbf{u}_4))(\mathbf{u}_4 - E(\mathbf{u}_4))^T\mathbf{A}_{44}^{-1}\mathbf{A}_{14}^T\right] \\ &= \mathbf{A}_{11} - \mathbf{A}_{14}\mathbf{A}_{44}^{-1}\text{Cov}(\mathbf{u}_4, \mathbf{u}_4)\mathbf{A}_{44}^{-T}\mathbf{A}_{14}^T \\ &= \mathbf{A}_{11} - \mathbf{A}_{14}\mathbf{A}_{44}^{-1}\mathbf{A}_{44}\mathbf{A}_{44}^{-T}\mathbf{A}_{14}^T \\ &= \mathbf{A}_{11} - \mathbf{A}_{14}\mathbf{A}_{44}^{-1}\mathbf{A}_{41} \end{aligned} \quad (10)$$

in which, $\text{Cov}(\mathbf{u}_4, \mathbf{u}_4) = \text{Var}(\mathbf{u}_4) = \mathbf{G}\sigma_u^2 = E(\mathbf{A}_{44})\sigma_u^2$, in addition, $\mathbf{A}$ is a symmetric

matrix, so its partition matrices are all symmetric as well, $\mathbf{A}_{44}^{-T} = \mathbf{A}_{44}^{-1}$, $\mathbf{A}_{14}^T = \mathbf{A}_{41}$.

Therefore, because the individuals with subscript 1 have no genotypes and omics information, the variance depends on their pedigree relationships with genotyped and measured intermediate omics individuals. In this way, variances and covariances are:

$$\begin{aligned}
\text{Var}(\mathbf{u}_1) &= \text{Var}(\mathbf{A}_{14}\mathbf{A}_{44}^{-1}\mathbf{u}_4 + \boldsymbol{\varepsilon}) \\
&= \text{Var}(\mathbf{A}_{14}\mathbf{A}_{44}^{-1}\mathbf{u}_4) + \text{Var}(\boldsymbol{\varepsilon}) \\
&= \mathbf{A}_{14}\mathbf{A}_{44}^{-1}\mathbf{G}\mathbf{A}_{44}^{-1}\mathbf{A}_{41} + \mathbf{A}_{11} - \mathbf{A}_{14}\mathbf{A}_{44}^{-1}\mathbf{A}_{41} \\
&= \mathbf{A}_{11} + \mathbf{A}_{14}\mathbf{A}_{44}^{-1}(\mathbf{G} - \mathbf{A}_{44})\mathbf{A}_{44}^{-1}\mathbf{A}_{41}
\end{aligned} \tag{11}$$

$$\text{Var}(\mathbf{u}_4) = \mathbf{C} \tag{12}$$

$$\begin{aligned}
\text{Cov}(\mathbf{u}_1, \mathbf{u}_4) &= \text{Cov}(\mathbf{A}_{14}\mathbf{A}_{44}^{-1}\mathbf{u}_4 + \boldsymbol{\varepsilon}, \mathbf{u}_4) \\
&= \mathbf{A}_{14}\mathbf{A}_{44}^{-1}\text{Var}(\mathbf{u}_4) \\
&= \mathbf{A}_{14}\mathbf{A}_{44}^{-1}\mathbf{C}
\end{aligned} \tag{13}$$

Like this, we can infer the covariances and variances between individuals belongs to different groups.

For example, the joint distribution of breeding values of $\mathbf{u}_1$ and $\mathbf{u}_2$ is:

$$p(\mathbf{u}_1, \mathbf{u}_2) = p(\mathbf{u}_2)p(\mathbf{u}_1|\mathbf{u}_2) \tag{14}$$

$$p(\mathbf{u}_4) = N(0, \mathbf{G}) \tag{15}$$

The conditional distribution of breeding values for Group 1 and Group 2 is:

$$p(\mathbf{u}_1|\mathbf{u}_2) = N(\mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{u}_2, \mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}) \tag{16}$$

$$\begin{aligned}
\mathbf{u}_1 &= E(\mathbf{u}_1|\mathbf{u}_2) + \boldsymbol{\varepsilon} \\
&= E(\mathbf{u}_1) + \mathbf{A}_{12}\mathbf{A}_{22}^{-1}[\mathbf{u}_2 - E(\mathbf{u}_2)] + \boldsymbol{\varepsilon} \\
&= \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{u}_2 + \boldsymbol{\varepsilon}
\end{aligned} \tag{17}$$

$$\begin{aligned}
\text{Var}(\boldsymbol{\varepsilon}) &= \text{Var}(\mathbf{u}_1 - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{u}_2) \\
&= \text{Var}(\mathbf{u}_1) - \text{Var}(\mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{u}_2) \\
&= \mathbf{A}_{11} - E\left[(\mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{u}_2 - E[\mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{u}_2])(\mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{u}_2 - E[\mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{u}_2])^T\right] \\
&= \mathbf{A}_{11} - E\left[\mathbf{A}_{12}\mathbf{A}_{22}^{-1}(\mathbf{u}_2 - E(\mathbf{u}_2))(\mathbf{u}_2 - E(\mathbf{u}_2))^T\mathbf{A}_{22}^{-T}\mathbf{A}_{12}^T\right] \\
&= \mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\text{Cov}(\mathbf{u}_2, \mathbf{u}_2)\mathbf{A}_{22}^{-T}\mathbf{A}_{12}^T \\
&= \mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{22}\mathbf{A}_{22}^{-T}\mathbf{A}_{12}^T \\
&= \mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}
\end{aligned} \tag{18}$$

$$\begin{aligned}
\text{Var}(\mathbf{u}_1) &= \text{Var}(\mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{u}_2 + \boldsymbol{\varepsilon}) \\
&= \text{Var}(\mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{u}_2) + \text{Var}(\boldsymbol{\varepsilon}) \\
&= \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{G}\mathbf{A}_{22}^{-1}\mathbf{A}_{21} + \mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21} \\
&= \mathbf{A}_{11} + \mathbf{A}_{12}\mathbf{A}_{22}^{-1}(\mathbf{G} - \mathbf{A}_{22})\mathbf{A}_{22}^{-1}\mathbf{A}_{21}
\end{aligned} \tag{19}$$

$$\text{Var}(\mathbf{u}_2) = \text{Var}(\mathbf{Zg}) = \mathbf{G} \tag{20}$$

$$\begin{aligned}
\text{Cov}(\mathbf{u}_1, \mathbf{u}_2) &= \text{Cov}(\mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{u}_2 + \boldsymbol{\varepsilon}, \mathbf{u}_2) \\
&= \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\text{Var}(\mathbf{u}_2) \\
&= \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{G}
\end{aligned} \tag{21}$$

In addition, the joint distribution of breeding values of $\mathbf{u}_1$ and $\mathbf{u}_3$ is:

$$p(\mathbf{u}_1, \mathbf{u}_3) = p(\mathbf{u}_3)p(\mathbf{u}_1|\mathbf{u}_3) \tag{22}$$

$$p(\mathbf{u}_3) = N(\mathbf{0}, \mathbf{M}) \tag{23}$$

The conditional distribution of breeding values for Group 1 and Group 3 is:

$$p(\mathbf{u}_1|\mathbf{u}_3) = N(\mathbf{A}_{13}\mathbf{A}_{33}^{-1}\mathbf{u}_3, \mathbf{A}_{11} - \mathbf{A}_{13}\mathbf{A}_{33}^{-1}\mathbf{A}_{31}) \tag{24}$$

$$\begin{aligned}
\mathbf{u}_1 &= E(\mathbf{u}_1|\mathbf{u}_3) + \boldsymbol{\varepsilon} \\
&= E(\mathbf{u}_1) + \mathbf{A}_{13}\mathbf{A}_{33}^{-1}[\mathbf{u}_3 - E(\mathbf{u}_3)] + \boldsymbol{\varepsilon} \\
&= \mathbf{A}_{13}\mathbf{A}_{33}^{-1}\mathbf{u}_3 + \boldsymbol{\varepsilon}
\end{aligned} \tag{25}$$

$$\begin{aligned}
\text{Var}(\boldsymbol{\varepsilon}) &= \text{Var}(\mathbf{u}_1 - \mathbf{A}_{13}\mathbf{A}_{33}^{-1}\mathbf{u}_3) \\
&= \text{Var}(\mathbf{u}_1) - \text{Var}(\mathbf{A}_{13}\mathbf{A}_{33}^{-1}\mathbf{u}_3) \\
&= \mathbf{A}_{11} - E\left[(\mathbf{A}_{13}\mathbf{A}_{33}^{-1}\mathbf{u}_3 - E[\mathbf{A}_{13}\mathbf{A}_{33}^{-1}\mathbf{u}_3])(\mathbf{A}_{13}\mathbf{A}_{33}^{-1}\mathbf{u}_3 - E[\mathbf{A}_{13}\mathbf{A}_{33}^{-1}\mathbf{u}_3])^T\right] \\
&= \mathbf{A}_{11} - E\left[\mathbf{A}_{13}\mathbf{A}_{33}^{-1}(\mathbf{u}_3 - E(\mathbf{u}_3))(\mathbf{u}_3 - E(\mathbf{u}_3))^T \mathbf{A}_{33}^{-T} \mathbf{A}_{13}^T\right] \\
&= \mathbf{A}_{11} - \mathbf{A}_{13}\mathbf{A}_{33}^{-1}\text{Cov}(\mathbf{u}_3, \mathbf{u}_3)\mathbf{A}_{33}^{-T} \mathbf{A}_{13}^T \\
&= \mathbf{A}_{11} - \mathbf{A}_{13}\mathbf{A}_{33}^{-1}\mathbf{A}_{33}\mathbf{A}_{33}^{-T} \mathbf{A}_{13}^T \\
&= \mathbf{A}_{11} - \mathbf{A}_{13}\mathbf{A}_{33}^{-1}\mathbf{A}_{31}
\end{aligned} \tag{26}$$

$$\begin{aligned}
\text{Var}(\mathbf{u}_1) &= \text{Var}(\mathbf{A}_{13}\mathbf{A}_{33}^{-1}\mathbf{u}_3 + \boldsymbol{\varepsilon}) \\
&= \text{Var}(\mathbf{A}_{13}\mathbf{A}_{33}^{-1}\mathbf{u}_3) + \text{Var}(\boldsymbol{\varepsilon}) \\
&= \mathbf{A}_{13}\mathbf{A}_{33}^{-1}\mathbf{G}\mathbf{A}_{33}^{-1}\mathbf{A}_{31} + \mathbf{A}_{11} - \mathbf{A}_{13}\mathbf{A}_{33}^{-1}\mathbf{A}_{31} \\
&= \mathbf{A}_{11} + \mathbf{A}_{13}\mathbf{A}_{33}^{-1}(\mathbf{G} - \mathbf{A}_{33})\mathbf{A}_{33}^{-1}\mathbf{A}_{31}
\end{aligned} \tag{27}$$

$$\text{Var}(\mathbf{u}_3) = \text{Var}(\mathbf{Zm}) = \mathbf{M} \tag{28}$$

$$\begin{aligned}
\text{Cov}(\mathbf{u}_1, \mathbf{u}_3) &= \text{Cov}(\mathbf{A}_{13}\mathbf{A}_{33}^{-1}\mathbf{u}_3 + \boldsymbol{\varepsilon}, \mathbf{u}_3) \\
&= \mathbf{A}_{13}\mathbf{A}_{33}^{-1}\text{Var}(\mathbf{u}_3) \\
&= \mathbf{A}_{13}\mathbf{A}_{33}^{-1}\mathbf{M}
\end{aligned} \tag{29}$$

Additionally, the joint distribution of breeding values of $\mathbf{u}_2$ and $\mathbf{u}_3$ is:

$$p(\mathbf{u}_2, \mathbf{u}_3) = p(\mathbf{u}_3)p(\mathbf{u}_2|\mathbf{u}_3) \tag{30}$$

$$p(\mathbf{u}_3) = N(\mathbf{0}, \mathbf{M}) \tag{31}$$

The conditional distribution of breeding values for Group 2 and Group 3 is:

$$p(\mathbf{u}_2|\mathbf{u}_3) = N(\mathbf{A}_{23}\mathbf{A}_{33}^{-1}\mathbf{u}_3, \mathbf{A}_{22} - \mathbf{A}_{23}\mathbf{A}_{33}^{-1}\mathbf{A}_{32}) \quad (32)$$

$$\begin{aligned} \mathbf{u}_2 &= E(\mathbf{u}_2|\mathbf{u}_3) + \boldsymbol{\varepsilon} \\ &= E(\mathbf{u}_2) + \mathbf{A}_{23}\mathbf{A}_{33}^{-1}[\mathbf{u}_3 - E(\mathbf{u}_3)] + \boldsymbol{\varepsilon} \\ &= \mathbf{A}_{23}\mathbf{A}_{33}^{-1}\mathbf{u}_3 + \boldsymbol{\varepsilon} \end{aligned} \quad (33)$$

$$\begin{aligned} \text{Var}(\boldsymbol{\varepsilon}) &= \text{Var}(\mathbf{u}_2 - \mathbf{A}_{23}\mathbf{A}_{33}^{-1}\mathbf{u}_3) \\ &= \text{Var}(\mathbf{u}_2) - \text{Var}(\mathbf{A}_{23}\mathbf{A}_{33}^{-1}\mathbf{u}_3) \\ &= \mathbf{A}_{22} - E[(\mathbf{A}_{23}\mathbf{A}_{33}^{-1}\mathbf{u}_3 - E[\mathbf{A}_{23}\mathbf{A}_{33}^{-1}\mathbf{u}_3])(\mathbf{A}_{23}\mathbf{A}_{33}^{-1}\mathbf{u}_3 - E[\mathbf{A}_{23}\mathbf{A}_{33}^{-1}\mathbf{u}_3])^T] \\ &= \mathbf{A}_{22} - E[\mathbf{A}_{23}\mathbf{A}_{33}^{-1}(\mathbf{u}_3 - E(\mathbf{u}_3))(\mathbf{u}_3 - E(\mathbf{u}_3))^T \mathbf{A}_{33}^{-T} \mathbf{A}_{23}^T] \\ &= \mathbf{A}_{22} - \mathbf{A}_{23}\mathbf{A}_{33}^{-1} \text{Cov}(\mathbf{u}_3, \mathbf{u}_3) \mathbf{A}_{33}^{-T} \mathbf{A}_{23}^T \\ &= \mathbf{A}_{22} - \mathbf{A}_{23}\mathbf{A}_{33}^{-1} \mathbf{A}_{33} \mathbf{A}_{33}^{-T} \mathbf{A}_{23}^T \\ &= \mathbf{A}_{22} - \mathbf{A}_{23}\mathbf{A}_{33}^{-1} \mathbf{A}_{32} \end{aligned} \quad (34)$$

$$\begin{aligned} \text{Var}(\mathbf{u}_2) &= \text{Var}(\mathbf{A}_{23}\mathbf{A}_{33}^{-1}\mathbf{u}_3 + \boldsymbol{\varepsilon}) \\ &= \text{Var}(\mathbf{A}_{23}\mathbf{A}_{33}^{-1}\mathbf{u}_3) + \text{Var}(\boldsymbol{\varepsilon}) \\ &= \mathbf{A}_{23}\mathbf{A}_{33}^{-1} \mathbf{G} \mathbf{A}_{33}^{-1} \mathbf{A}_{32} + \mathbf{A}_{22} - \mathbf{A}_{23}\mathbf{A}_{33}^{-1} \mathbf{A}_{32} \\ &= \mathbf{A}_{22} + \mathbf{A}_{23}\mathbf{A}_{33}^{-1}(\mathbf{G} - \mathbf{A}_{33})\mathbf{A}_{33}^{-1} \mathbf{A}_{32} \end{aligned} \quad (35)$$

$$\text{Var}(\mathbf{u}_3) = \text{Var}(\mathbf{Zm}) = \mathbf{M} \quad (36)$$

$$\begin{aligned} \text{Cov}(\mathbf{u}_2, \mathbf{u}_3) &= \text{Cov}(\mathbf{A}_{23}\mathbf{A}_{33}^{-1}\mathbf{u}_3 + \boldsymbol{\varepsilon}, \mathbf{u}_3) \\ &= \mathbf{A}_{23}\mathbf{A}_{33}^{-1} \text{Var}(\mathbf{u}_3) \\ &= \mathbf{A}_{23}\mathbf{A}_{33}^{-1} \mathbf{M} \end{aligned} \quad (37)$$

Finally, the matrix containing the joint relationships is given by:

$$\begin{aligned} \mathbf{FS} &= \begin{bmatrix} \mathbf{H}_{11} & \mathbf{H}_{12} & \mathbf{H}_{13} & \mathbf{H}_{14} \\ \mathbf{H}_{21} & \mathbf{H}_{22} & \mathbf{H}_{23} & \mathbf{H}_{24} \\ \mathbf{H}_{31} & \mathbf{H}_{32} & \mathbf{H}_{33} & \mathbf{H}_{34} \\ \mathbf{H}_{41} & \mathbf{H}_{42} & \mathbf{H}_{43} & \mathbf{H}_{44} \end{bmatrix} \\ &= \begin{bmatrix} \text{Var}(\mathbf{u}_1) & \text{Cov}(\mathbf{u}_1, \mathbf{u}_2) & \text{Cov}(\mathbf{u}_1, \mathbf{u}_3) & \text{Cov}(\mathbf{u}_1, \mathbf{u}_4) \\ \text{Cov}(\mathbf{u}_2, \mathbf{u}_1) & \text{Var}(\mathbf{u}_2) & \text{Cov}(\mathbf{u}_2, \mathbf{u}_3) & \text{Cov}(\mathbf{u}_2, \mathbf{u}_4) \\ \text{Cov}(\mathbf{u}_3, \mathbf{u}_1) & \text{Cov}(\mathbf{u}_3, \mathbf{u}_2) & \text{Var}(\mathbf{u}_3) & \text{Cov}(\mathbf{u}_3, \mathbf{u}_4) \\ \text{Cov}(\mathbf{u}_4, \mathbf{u}_1) & \text{Cov}(\mathbf{u}_4, \mathbf{u}_2) & \text{Cov}(\mathbf{u}_4, \mathbf{u}_3) & \text{Var}(\mathbf{u}_4) \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{A}_{11} + \mathbf{A}_{14}\mathbf{A}_{44}^{-1}(\mathbf{C} - \mathbf{A}_{44})\mathbf{A}_{44}^{-1}\mathbf{A}_{41} & \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{G} & \mathbf{A}_{13}\mathbf{A}_{33}^{-1}\mathbf{M} & \mathbf{A}_{14}\mathbf{A}_{44}^{-1}\mathbf{C} \\ \mathbf{G}_{22}\mathbf{A}_{22}^{-1}\mathbf{A}_{21} & \mathbf{G}_{22} & \mathbf{A}_{23}\mathbf{A}_{33}^{-1}\mathbf{M} & \mathbf{A}_{24}\mathbf{A}_{44}^{-1}\mathbf{C} \\ \mathbf{M}\mathbf{A}_{33}^{-1}\mathbf{A}_{31} & \mathbf{M}\mathbf{A}_{33}^{-1}\mathbf{A}_{32} & \mathbf{M}_{33} & \mathbf{A}_{34}\mathbf{A}_{44}^{-1}\mathbf{C} \\ \mathbf{C}\mathbf{A}_{44}^{-1}\mathbf{A}_{41} & \mathbf{C}\mathbf{A}_{44}^{-1}\mathbf{A}_{42} & \mathbf{C}\mathbf{A}_{44}^{-1}\mathbf{A}_{43} & \mathbf{C} \end{bmatrix} \end{aligned} \quad (38)$$

$$\text{Where } \mathbf{C} = \alpha \mathbf{M}_{44} + \beta \mathbf{G}_{44} + (1 - \alpha - \beta) \mathbf{A}_{44}$$

The matrix in question serves as an advanced rendition of the conventional $\mathbf{H}$ matrix utilized within the ssGBLUP methodology. A critical distinction must be highlighted: this $\mathbf{FS}$ matrix presents a divergence from the $\mathbf{H}$ matrix in the traditional model. Should there be an absence of data pertaining to omics expression levels, the genotyped individuals are encapsulated within the animal subset denoted by subscript 2. Under such circumstances, the $\mathbf{FS}$ matrix reverts to its conventional structure, the $\mathbf{H}$ matrix, as employed in standard single-step genetic evaluations.

References cited in this Additional file have been included in the main article reference list.

Supplementary analysis

The Cancer Genome Atlas (TCGA) dataset

Data from TCGA-OV and TCGA-STAD were collected. While full genetic (SNP) data is restricted, we leveraged publicly available mRNA, miRNA, and lncRNA data (Level 3) using the TCGAbiolinks package. The clinical variables including age, tumor stage, vital status (alive or deceased) and days to death. Gene expression levels were calculated as fragments per kilobase of transcript per million mapped reads (FPKM). Genes with FPKM < 0.1 in more than 95% of the samples were discarded. After these quality control steps, there's 211 samples within STAD and 265 samples within OV datasets, respectively.

We compared FSBLUP against a series of baseline models: 1) Single-omics models (M1): Predictions based solely on the similarity kernel from a single omics layer (lncRNA, mRNA, or miRNA), 2) Additive multi-omics models (M2): Predictions based on the unweighted sum of two or three similarity kernels (e.g., $K_{\text{mRNA}} + K_{\text{miRNA}}$ for a two-kernel model; $K_{\text{mRNA}} + K_{\text{miRNA}} + K_{\text{lncRNA}}$ for a three-kernel model), and 3) FSBLUP:, which integrates all three available omics kernels (mRNA, miRNA, lncRNA) through machine learning-optimized weighting. Prediction accuracy was assessed by calculating the correlation coefficient between predicted value and response variable using five-fold cross-validation repeated 20 times.