## S2 Note. FLM-based association analysis under different ratios between $\boldsymbol{m}, \boldsymbol{K}_{\boldsymbol{G}}$ and $\boldsymbol{K}_{\boldsymbol{\beta}}$

## Necessary conditions for performance of FLM-based method

In the framework of a functional linear model (FLM), we fix the number of basis functions ( $K_{G}$ to estimate genetic variant functions (GVF) and $K_{\beta}$ to estimate beta-smooth function (BSF)) depending on the function basis type which can be different for GVF and BSF.

To perform association analysis using Model (5), we must estimate unknown betas, $\beta_{F}$, solving the equation

$$
\begin{equation*}
\beta_{F}=\left(W^{T} G^{T} \Omega^{-1} G W\right)^{-1} W^{T} G^{T} \Omega^{-1}(y-X \alpha), \tag{S1}
\end{equation*}
$$

where matrices $\Omega, G$, and $W$ have dimensions $n \times n, n \times m$ and $m \times K_{\beta}$, respectively. Inverting the $K_{\beta} \times K_{\beta}$ square matrix $W^{T} G^{T} \Omega^{-1} G W$ in formula (S1) is possible if $n \geq m \geq K_{\beta}$. Therefore to unambiguously evaluate $\beta_{F}$, we must restrict the number of basis functions, $K_{\beta}$, as

$$
K_{\beta} \leq m .
$$

To introduce additional, stricter restrictions preventing an over-parameterization in Model (5), we present a matrix $W$ of dimension $m \times K_{\beta}$ as a product of two matrices $W_{1}$ and $W_{2}$ :

$$
\begin{equation*}
W_{1}=\Phi\left(\Phi^{\mathrm{T}} \Phi\right)^{-1} \tag{S2}
\end{equation*}
$$

of dimension $m \times K_{G}$, and

$$
\begin{equation*}
W_{2}=\int_{0}^{1} \phi(t) \psi^{T}(t) d t \tag{S3}
\end{equation*}
$$

of dimension $K_{G} \times K_{\beta}$.
Then the matrix $W^{T} G^{T} \Omega^{-1} G W$ in (S1) is transformed into $W_{2}^{T} W_{1}^{T} G^{T} \Omega^{-1} G W_{1} W_{2}$. The additional restrictions must be imposed on dimensions of the matrices $W_{1}$ and $W_{2}$. Inverting the $K_{\beta} \times K_{\beta}$ square matrix $W_{2}{ }^{T} W_{1}^{T} G^{T} \Omega^{-1} G W_{1} W_{2}$ is possible if $n \geq m \geq K_{G} \geq K_{\beta}$. As a result, we must firstly restrict the number of basis functions for smoothing the genotypes, $K_{G}$, and then for smoothing the betas, $K_{\beta}$ :

$$
K_{G} \leq m \text { and } K_{\beta} \leq K_{G},
$$

respectively. Therefore, if the declared number of the basis functions, $K_{G}$, used for GVF is more than $m, K_{G}$ has to be reduced to $m$. Next, if the declared number of basis functions, $K_{\beta}$, used for BSF is more than $K_{G}, K_{\beta}$ has to be reduced to $K_{G}$.

## Equivalence of several functional linear models

In general, the functional linear models under consideration are described as

$$
\begin{equation*}
y=X \alpha+G W \beta_{F}+h+e, \tag{S4}
\end{equation*}
$$

and an $F$-test statistic for them is calculated as

$$
F=\frac{\left(R S S_{0}-R S S_{1}\right) / K_{\beta}}{R S S_{1} /\left(n-K_{\beta}-1\right)} .
$$

Here $\quad R S S_{0}=(y-X \alpha)^{T} \Omega^{-1}(y-X \alpha), \quad R S S_{1}=(y-X \alpha)^{T} P^{T} \Omega^{-1} P(y-X \alpha), \quad$ and $\quad P=I-$ $G W\left(W^{T} G^{T} \Omega^{-1} G W\right)^{-1} W^{T} G^{T} \Omega^{-1}$.

The difference between the models with and without the smoothing on the GVF is in the way the matrix $W$ is constructed. For all models using both the GVF and the BSF (i.e., F-F, B-B, F-B and B-F models), the matrix $W$ is formed as product of two matrices $W_{1}$ and $W_{2}$, defined by expressions (S2) and (S3). The matrix $W_{1}$ of dimension $m \times K_{G}$ participates in the genotypes smoothing on base of one function basis, and the matrix $W_{2}$ of dimension $K_{G} \times K_{\beta}$ participates in the betas smoothing (under condition of the GVF smoothing), using two function bases. For the models which do not use the GVF (i.e., 0-F and 0-B models), the matrix $W$ of dimension $m \times K_{\beta}$ is intended to smooth only the betas on the base of only one function basis and constructed analogously to the $m \times K_{G}$ matrix $\Phi$ in Model (5).

For the models with both the GVF and BSF, the matrix $P$ looks as:

$$
\begin{equation*}
P=I-G W_{1} W_{2}\left(W_{2}^{T} W_{1}^{T} G^{T} \Omega^{-1} G W_{1} W_{2}\right)^{-1} W_{2}^{T} W_{1}^{T} G^{T} \Omega^{-1} . \tag{S5}
\end{equation*}
$$

If $m>K_{G}$ and $K_{G}=K_{\beta}$, the matrix $W_{2}$ is invertible, and hence it can be canceled in the expression for $P$ (S5). Moreover, decomposing $W_{1}$ into $\Phi$ and ( $\left.\Phi^{\mathrm{T}} \Phi\right)^{-1}$ (see formula (S2)) allows us also to cancel the matrix $\left(\Phi^{\mathrm{T}} \Phi\right)^{-1}$ in the expression for $P(\mathrm{~S} 5)$. As a result, the matrix $W$ is reduced to the matrix $\Phi$. Therefore, the models within groups ( $0-\mathrm{F}, \mathrm{F}-\mathrm{F}$, and F-B) and ( $0-\mathrm{B}, \mathrm{B}-\mathrm{B}$ and B-F) do not differ from each other because the matrix $\Phi$ for the models with the GVF and the matrix $W$ for the models without the GVF are identical.

If $m \geq K_{G}$ and $K_{G}>K_{\beta}$, the matrix $W_{2}$ is not invertible (because it is not square) and it cannot be canceled in the expression for $P$ (S5). As a result, all the models are different because matrices $W$ in the models with both the GVF and the BSF and the models with only the GVF are constructed differently.

In our study, we fixed the number of basis functions as 25 for Fourier series and as 15 for B-splines. Therefore:

- The B-B model is equivalent to the $0-\mathrm{B}$ model, and the $\mathrm{F}-\mathrm{F}$ model is equivalent to the $0-\mathrm{F}$ model.
- The F-B model is equivalent to none of the used models.
- In the B-F model, $K_{G}$ (originally declared as 25 ) is automatically reduced to 15 to meet condition $K_{G} \geq K_{\beta}$. This model is equivalent to the models $0-\mathrm{B}$ and B-B.

In summary, the set of all six models used in our study is reduced to the set of three models being different from each other: 0-B (identical to B-B and B-F), 0-F (identical to F-F) and F-B.

## Analysis of regions with small number of genetic variants

For the situations, when the number of genetic variants is small (smaller than the number of basis functions), to avoid model over-parameterization, we imposed restrictions on the number of basis functions, reducing $K_{G}$ and $K_{\beta}$ to the number of variants in a region of interest. As a result, if $K_{G}$ and $K_{\beta}$ become equal to $m$, the functional linear model (5), described by the equation (S4), is reduced to a more simple linear mixed model (1), described by the equation

$$
y=X \alpha+G \beta+h+e,
$$

because the matrix $W$ becomes invertible and it can be canceled in the expression for $P$ (S5). Therefore, the benefit of the proposed method compared to simple approach based on linear mixed model exists only for regions with the large number of variants.

