Supplementary Material

Claim: *Correlation between B(i), B(j), the i^{th} and the j^{th} components of B is negative when i,j are alternative features, which implies the assumption that $f(X) = f(X_{swap(i,j)})$, and X and $X_{swap(i,j)}$ follows the same distribution. $X_{swap(i,j)}$ is the matrix from exchanging the i^{th} column and j^{th} column of X.*

Proof:

Denote X as the feature vector, a D dimensional random vector, and Y is the responding variable.
First we assume the ground true model is represented by the following formula.

$$Y = f(X)$$

When two genetic variants X_i, X_j are a pair of alternative genetic variants for a responding variable, we assume that exchanging their values will not change the responding variable. Formally, it is defined as

$$f(X_{\mathrm{swap}(i,j)}) = f(X)\cdot \tag{2}$$

We also assume $X_{\mathrm{swap}(i,j)}$ and X follows the same distribution.
The optimal solution of a sparse model can be written as B.

$$B = \mathrm{argmin}_B[f(X) - X^T B]^2 + \lambda|B|_1\cdot \tag{3}$$

So the solution will satisfy the following Karush–Kuhn–Tucker condition.
$\nabla_B[(Y - XB)^2 + \lambda|B|_1] = 0$ which is

$$X[X^T B - f(X)] = \nabla_B[\lambda|B|_1]. \tag{4}$$

Since $X^T B = X^T_{\mathrm{swap}(i,j)} B_{\mathrm{swap}(i,j)}$ and $f(X_{\mathrm{swap}(i,j)}) = f(X)$, we have the following,

$$X\,[X^T_{\mathrm{swap}(i,j)} B_{\mathrm{swap}(i,j)} - f(X_{\mathrm{swap}(i,j)})] = \nabla_B[\lambda|B|_1].$$

We can also swap the i^{th} and j^{th} column of each side of the equation.
$X_{\mathrm{swap}(i,j)}\,[X^T_{\mathrm{swap}(i,j)} B_{\mathrm{swap}(i,j)} - f(X_{\mathrm{swap}(i,j)})] = \nabla_{B_{\mathrm{swap}(i,j)}}[\lambda|B_{\mathrm{swap}(i,j)}|_1]$, which is

$$X\,[X^T B_{\mathrm{swap}(i,j)} - X_{\mathrm{swap}(i,j)} f(X_{\mathrm{swap}(i,j)})] = \nabla_{B_{\mathrm{swap}(i,j)}}[\lambda|B_{\mathrm{swap}(i,j)}|_1].$$

Then we compute the expectation of the left side over all X and get the following equation.

$$E[XX^T]B_{\mathrm{swap}(i,j)} - E\,[X_{\mathrm{swap}(i,j)} f(X_{\mathrm{swap}(i,j)})] = E[XX^T]B_{\mathrm{swap}(i,j)} - E[Xf(X)] = \nabla_{B_{\mathrm{swap}(i,j)}}[\lambda|B_{\mathrm{swap}(i,j,)}|_1]. \tag{5}$$

Therefore, $B_{\mathrm{swap}(i,j)}$ and B are solutions of the equation

$$E[XX^T]B - E[Xf(X)] = \nabla_B[\lambda|B|_1]. \tag{6}$$

Then from a linear convex combination of Equations (5) and (6), we can know $(1-a)B + aB_{swap(i,j)}$, $0 < a < 1$, is also an optimal solution of the problem in Equation (3).

Because $B(i) = B_{swap(i,j)}(j)$ and $B(j) = B_{swap(i,j)}(i)$, for all the model $(1-a)B + a\ B_{swap(i,j)}$, the i^{th} component is going from $B(i)$ to $B(j)$,while the j^{th} component is going from $B(j)$ to $B(i)$. Thus on the condition that there is a solution with $B(i) \neq B(j)$, the correlation coefficient of the i^{th} and the j^{th} component among all those models is -1. Q.E.D. ■