**Supplementary Materials**

**Supplementary Figure 1**. Sensitivity of the in the first stage of ProSGPV

Caption: Mean capture rates of the exact true model are surrounded by 95% Wald confidence intervals over 1000 simulations. Black is the ProSGPV algorithm implemented with , which is the that minimizes the generalized information criterion. Green is the ProSGPV algorithm implemented with from a uniform distribution of (0.8, 1.2), where is the that minimizes the cross validation error, and is the largest that yields a cross validation error that is within one standard error of the minimal cross validation error.

**Supplementary Figure 2**. Sensitivity of the null bound in ProSGPV

Caption: Different null bounds and their corresponding support recovery rates are compared in Logistic regression at three correlation levels. Medians are surrounded with first and third quartiles from 1000 simulations. Null bound choices include the original null bound , , , , and 0.

**Supplementary Figure 3.** Comparison of maximum likelihood fitted and Jeffreys prior penalized logistic regressions.

Caption: Average capture rates of the exact true model, average mean absolute errors (MAE), and average prediction area under the curve (AUC) in a separate test set are compared over 1000 simulations. For capture rates, means are surrounded by 95% Wald confidence intervals. For MAE and prediction AUC, medians are surrounded by first and third quartiles from the simulation.

**Supplementary Figure 4**. Comparison of Type I Error rate and Power for all algorithms

Caption: Average Type I Error rates and estimated power rates from 1000 simulations are compared for all algorithms. Average Type I Error rates are calculated as the average proportion of falsefully identified signal variables among all noise variables. Estimated power rates are calculated as the average proportions of correctly identified signal variables.

**Supplementary Figure 5.** Comparison of FDR and FNDR for all algorithms

Caption: Average false discovery proportions (pFDR) and false non-discovery proportions (pFNDR) are compared for all algorithms over 1000 simulations. pFDR is calculated as the proportion of identified “signal” variables that are indeed noise variables. pFNDR is calculated as the proportion of identified “noise” variables that are indeed signal variables.

**Supplementary Figure 6**. Comparison of a constant null bound and a data-dependent null bound

Caption: Support recovery rate, parameter estimation mean absolute error (MAE), and prediction area under the curve (AUC) are compared for the ProSGPV with a constant null bound and the ProSGPV with a generalized variance inflation factor adjusted null bound. Means (solid lines) and Wald 95% confidence intervals (shades) are compared for support recovery. Median (solid lines) and first and third quartiles (shades) MAEs are compared for parameter estimation. Median (solid lines) and first and third quartiles (shades) prediction AUC are compared for prediction using a separate test set.

**Supplementary Figure 7**. Comparison of computation time for all algorithms

Caption: Running time in seconds are compared for all algorithms over 1000 repetitions. For aesthetic reasons, data are capped at 1.2 seconds.

**Supplementary Figure 8**. Clustering and correlation pattern of the spine data

Caption: The positions of variables imply the clustering pattern. Color indicates the strength of pairwise correlation. Blue indicates a positive correlation and red indicates a negative correlation. The darker the color, the stronger the correlation.

**Supplementary Figure 9**. Comparison of sparsity of solutions from all algorithms

Caption: Density of model of each algorithm using the training data (70% of all data) are compared over 1000 repetitions. Most often selected models are also annotated with color.

**Supplementary Figure 10**. Comparison of prediction accuracy from all algorithms

Caption: Distribution of prediction area under the curve in the test set (30% of all data) of each algorithm are compared over 1000 repetitions.

**R Code**. R code to replicate results in the paper