**Supplementary Data**

**Table S1.** Pseudo code for mDAG

|  |  |
| --- | --- |
| **Input:** | Mixed data matrix for variables (nodes) . Each row of is a sample, each column is a variable (node). |
| **Output:** | by matrix , where represents the DAG.  indicates directed edge ; indicates node and are not connected. |
| **Step 1** | **1.1** For node from 1 to , run the -penalized GLM to identify the Markov Blanket of node , with the optimal tuning parameter being chosen by EBIC.  **1.2** Define a by matrix representing an undirected graph of the variables  (nodes) . For any pair of nodes and , if node is in the Markov blanket of node or node is in the Markov blanket of node , set ;  otherwise set . |
| **Step 2** | **2.1** For any pair of nodes and with ,  if node and are marginally independent based on the permutation  test, set .  **2.2** For any pair of nodes and with , let  be the set of nodes that could be common children or  descendants of and .  For all subsets , let      Test whether node and are conditional independent  given using the permutation test.  If they are conditionally independent, set |
| **Step 3** | **3.1** Let be an empty graph. Calculate its BIC score , where is the BIC score of node based on an empty graph.  **3.2** Perform Hill Climbing greedy search algorithm to add, reverse or delete edges.  Set count=0,  While (count<5)  For node from 1 to  For node from 1 to  Set  If and , set .  If reset  If ,  Case 1: set  If , reset  Case 2: set  If  reset  If , then count=count+1  else set |

**Table S2.** Summary of simulation scenarios.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Scenario | Sample size | Number of nodes | Percent of categorical nodes (%) | Number of edges |
| 1 | 100 | 100 | 10 | 100 |
| 2 | 100 | 100 | 20 | 100 |
| 3 | 1000 | 500 | 10 | 500 |
| 4 | 1000 | 500 | 20 | 500 |
| 5 | 100 | 100 | 10 | 500 |
| 6 | 100 | 100 | 20 | 500 |
| 7 | 1000 | 500 | 10 | 2500 |
| 8 | 1000 | 500 | 20 | 2500 |



Figure S1. Small-scale illustration of the mDAG algorithm. (a) True DAG; (b) Estimated MGM; (c) Estimated skeleton; (d) Estimated DAG.