**Supplementary File for: Biological network inference with GRASP: a Bayesian network structure learning method using adaptive sequential Monte Carlo**

Kaixian Yu1,\*, Zihan Cui1, Xin Sui1, Xing Qiu2 and Jinfeng Zhang1,\*

1Department of Statistics, Florida State University, Tallahassee, FL, 32306

2Department of Biostatistics and Computational Biology, University of Rochester, Rochester, NY 14624, USA

**Appendix**

**Preliminary of Bayesian Network**

Let us denote the set of variables (nodes) as , and the set of edges as where are directed edges in the graph. ­ is called a parent of and a child of . Thus a graph can be represented as .

**Definition 1:** For a set of nodes if the edges , then we say forms a directed path between and . If , then this directed path is called a cycle.

**Definition 2:** A directed acyclic graph (DAG) is a graph such that all edges in *E* are directed and there is no cycles in *G*.

We denote *P(****X****)* as a joint probability distribution over the random variables in ***X***, and as the set of parents of given DAG *G(****X****,****E****).*

**Property 1:** *P(****X****)* can be factorized over some G as

(1)

Now, we can define Bayesian network (BN) as follows:

**Definition 3:** The pair *(G,P)* is defined as a Bayesian network if *P* factorizes over *G*.

**Remark 1:** The factorization allowed the network to be locally trained, e.g. each can be trained independently, which saves a lot of computational time.

Unfortunately, the factorization is not uniquely defined, that is, for some *P* there exists at least two DAGs and that *P* factorizes over and .

**Definition 4:** ***Q****(P)* defines an equivalent class of *P*, if *P* factorizes over each DAG .

In this work, we focusing on estimating any instead of estimating every DAGs in .

Now let us define the conditional dependencies and independencies. We denote as and are conditionally independent given with respect to *P(****X****)*, and as and are conditionally dependent given with respect to .

**Definition 5**:

(2)

(3)

One of the most important assumptions we need to include is the faithfulness. To define the faithfulness, let us first define trail:

**Definition 6:** A set of nodes forms a trail in the graph if for every and , either or .

Before we can define an active trail, let us first define descendant. If is a descendant of then there is a directed path from and .

**Definition 7:** Let be a BN structure, forms a trail in and . The trail is active given if

• whenever there is a v-structure: , then or a descendant of in

• other nodes are not in

We need one last definition, d-separation, before we can define faithfulness.

**Definition 8:** In graph , for and , we say and are d-separated by , denoted as, if none of the trails between and is active given .

Now let us give the definition on the faithfulness,

**Definition 9:** is faithful to if for any **:**

(4)

Under the faithfulness assumption, the terms conditionally independence and d-separation are equivalent; thus, they were used interchangeably through the article.

**Definition 10** A score function is said to have score-equivalent property if do not distinguish among equivalent networks. That is, two Bayesian networks, and , are equivalent if and only if .

**Proof of Theorem 1**

Assume is a Bayesian network, and . Since factorizes on , implies and . Now let us assume there is an such that . Therefore, all path between and are blocked by . Since , all path between and are also blocked by ; thus, which contradicts with the assumption. Hence, if , then for any we have . Along with, we conclude and for any edge in .

**Remark 2:**  *can be written as*

In practice, a symmetric correction may be used, where if but then

C:\Users\kaikai\Dropbox\class_learning\kai\research\paper_in_progress\BN_smc\Draft_5_adaptive\adsmc.tif

**Figure S1**: Traditional SMC and adaptive SMC. This shows a local skeleton in some step of SMC. In a traditional SMC, it is likely the update sequence set to be , while this might limit the later update to only 2 configurations instead of 3. In adSMC, by selecting sequence based on current structure, we will sample first, which gives us more options on the local sctructure for , and this configuration will not affect the options for or ; thus, increased the diversity of samples.

地图的截图

描述已自动生成地图的截图

描述已自动生成

**Figure S2**: F1-score and recall over different methods with observation size 2000 and 5000.



**Figure S3**: The trend of the BIC scores of the learned networks as temperature increases. Observation sizes: (A) 1000, (B) 2000, (C) 5000

C:\Users\Kaixian\Dropbox\class_learning\kai\research\paper_in_progress\BN_smc\Draft_5_adaptive\adsmc_fixsmc.tif

**Figure S4:** The comparison of adSMC and traditional SMC. (A) Andes, (B) hepar2



**Figure S5**: The empirical distribution of BIC scores of all SMC samples. (A) alarm network, (B) win95pts network

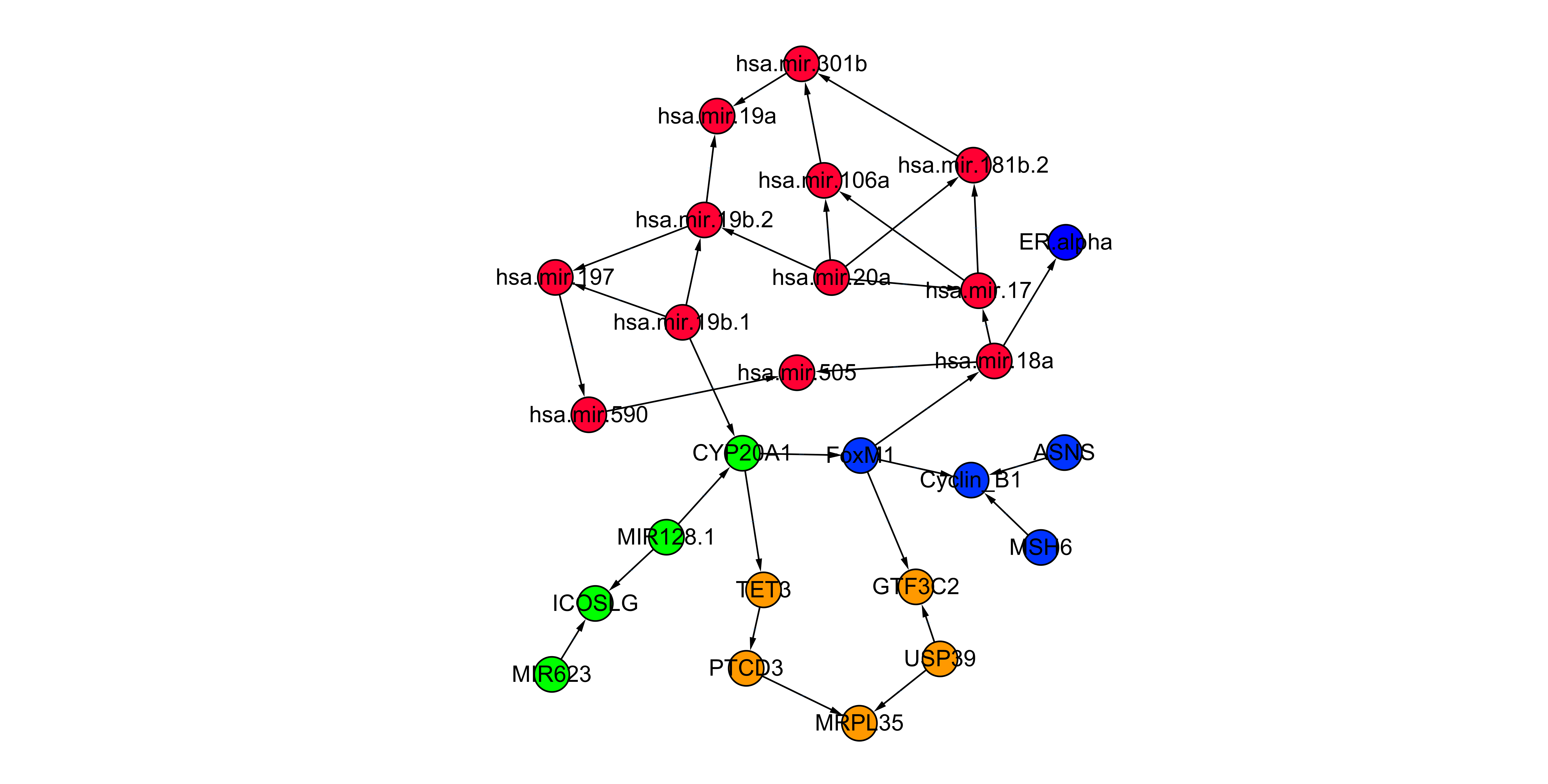
手机屏幕截图

描述已自动生成

手机屏幕截图

描述已自动生成

**Figure S6**: BIC scores of all methods on 7 benchmark networks with observation size 2000 and 5000.



**Figure S7**: The BN structure learned by bnlearn using multiple different genomic features which are highly correlated with the expression of LOC90784. Orange nodes: mRNA transcripts; Red nodes: microRNAs; Blue nodes: protein expressions; Green nodes: DNA methylations.