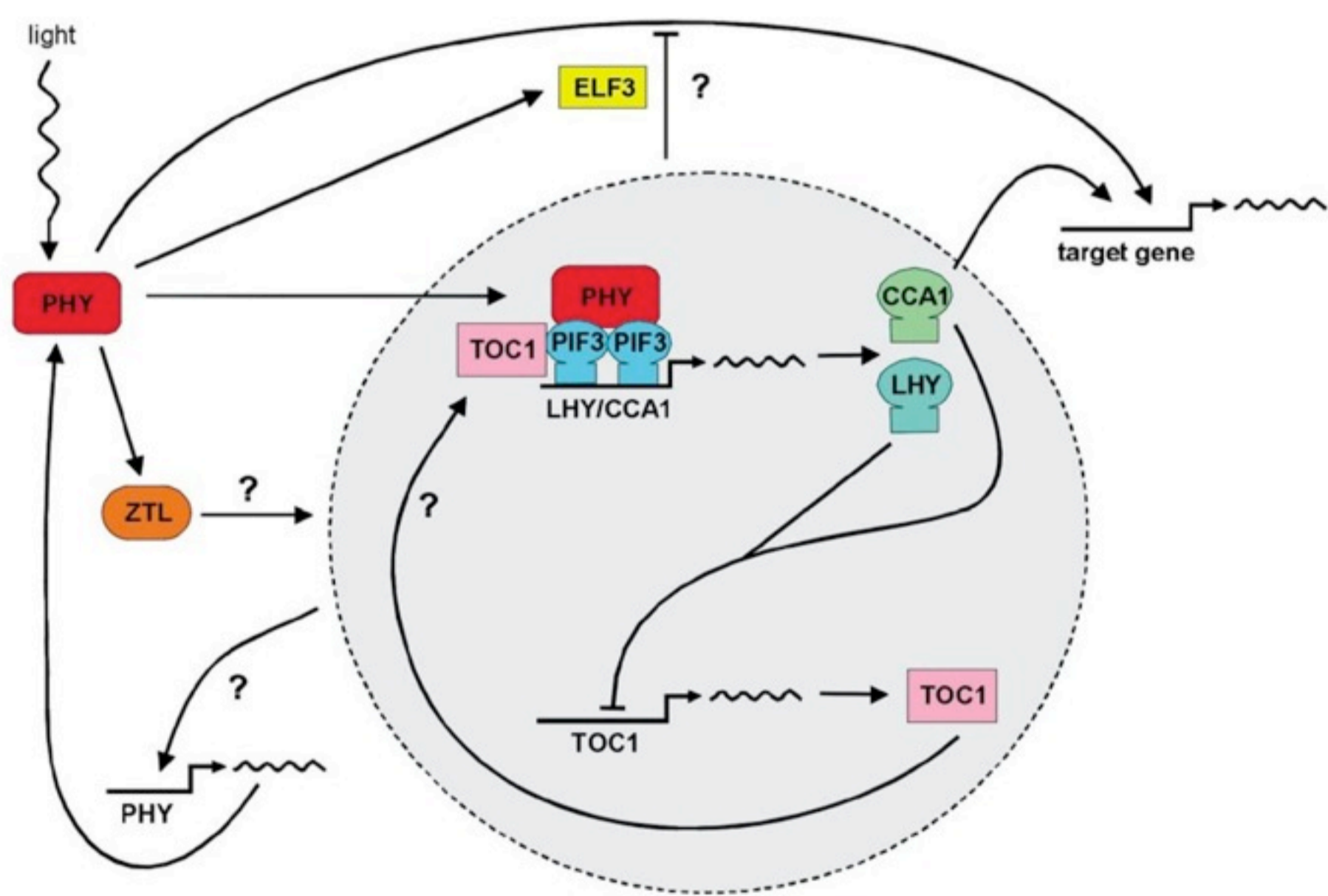


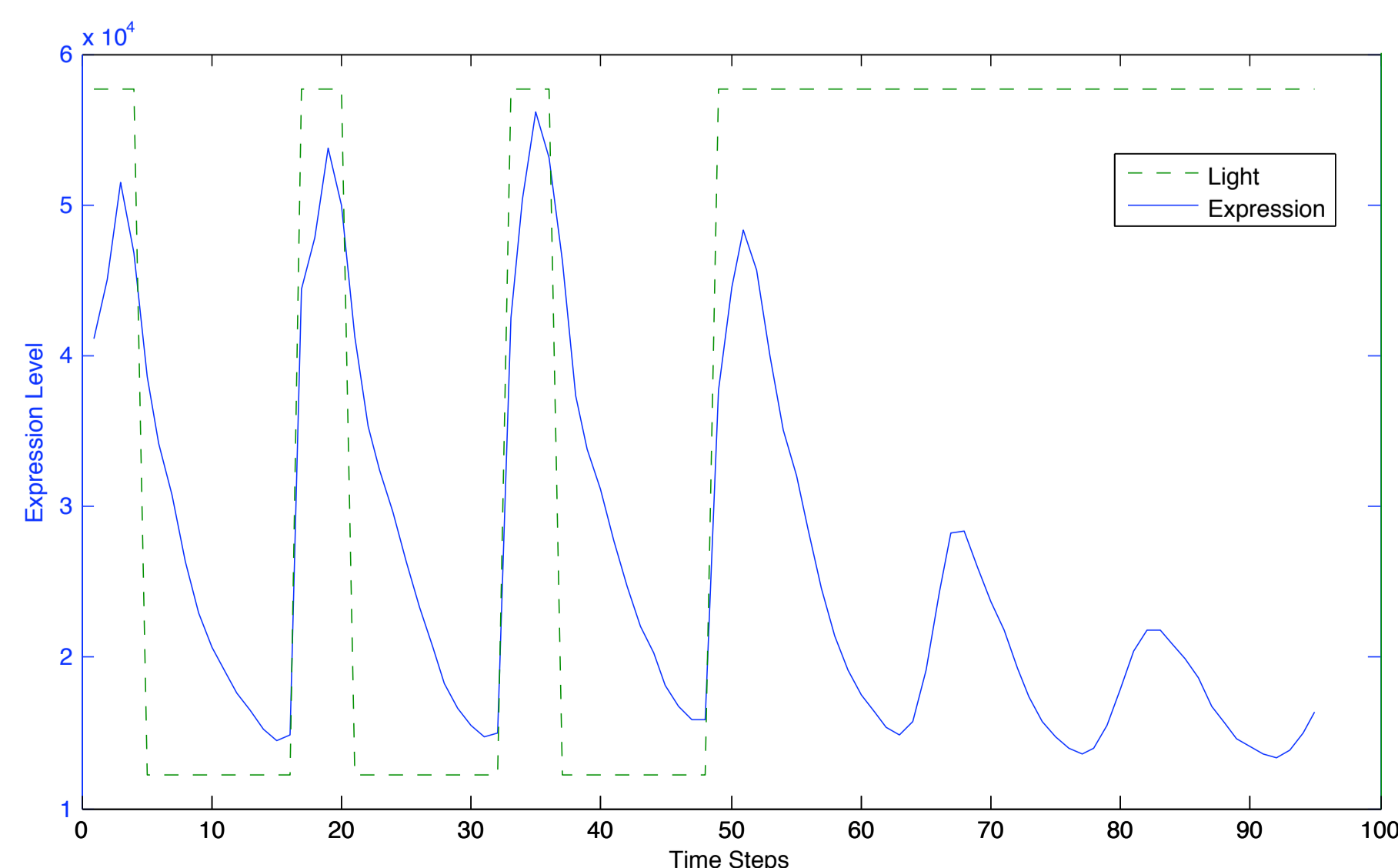
This poster shows how the circadian clock of *Arabidopsis thaliana* is modelled by fitting dynamic Bayesian networks to luminescence data gathered from experiments. This work differs from previous modelling attempts by using higher-order dynamic Bayesian networks to explicitly model the time lag between the various genes being expressed.

Gene Regulatory Networks

A gene regulatory network describes how genes interact with each other through expression of proteins. Such proteins act as transcription factors and encourage or suppress the expression of other genes. Parts of networks can have feedback. With external stimulation, these can act as a clock.



Certain genes display periodic behaviour, when periodically stimulated by light. However, this behaviour continues when the periodic stimulation is removed, with a gradual decay.



Bayesian Networks

A Bayesian networks is a representation of a joint probability distribution. It consists of a DAG structure and conditional probability distributions associated with each node on the DAG. If given a DAG G and joint distribution P ,

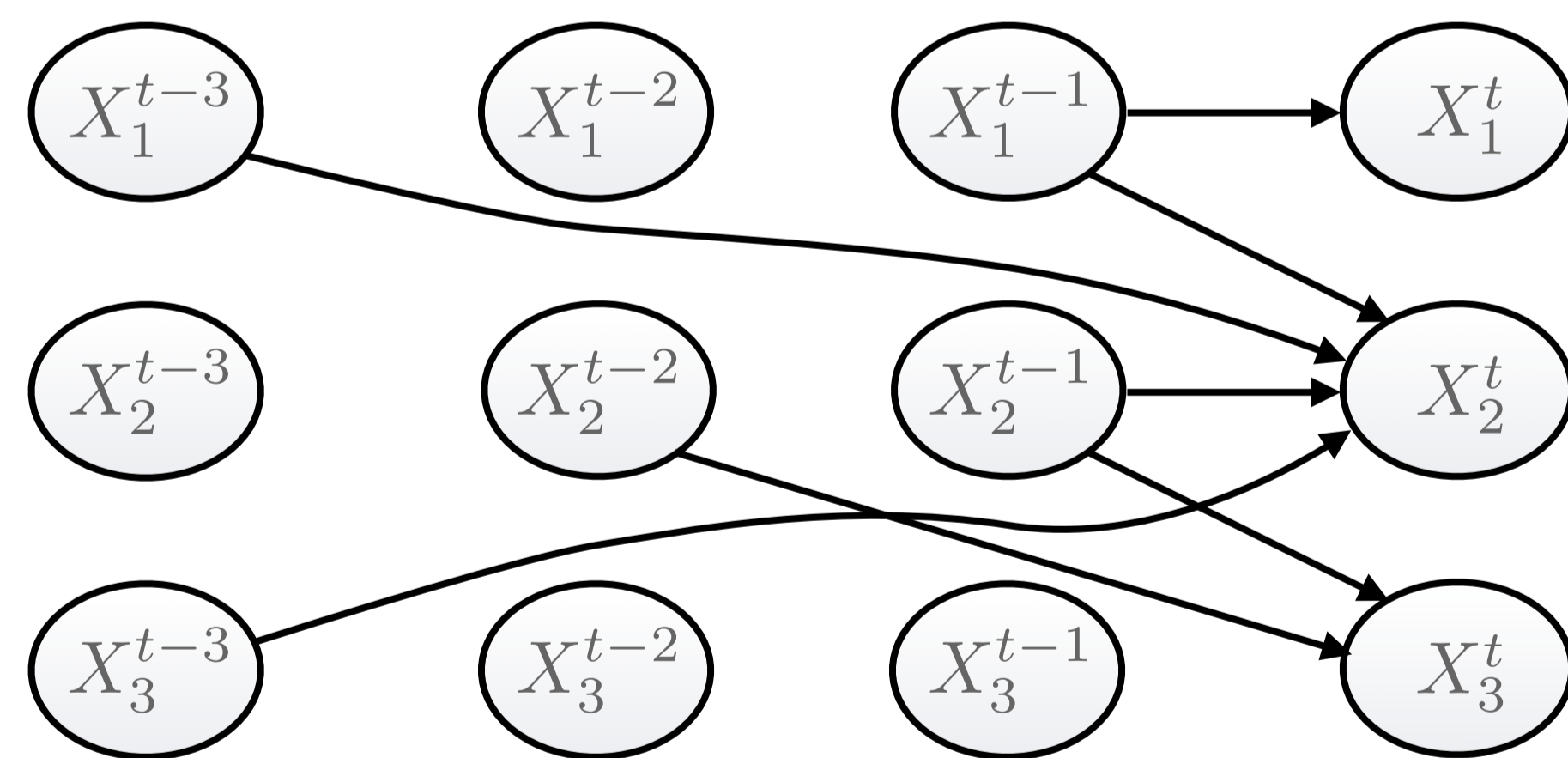
$$X \perp\!\!\!\perp_P ND(X) | Pa(X)$$

then G satisfies the Markov condition with P , and the joint is given by the product of the conditionals

$$P(X_1, X_2, \dots, X_n) = \prod_{n=1}^N P(X_n | Pa(X_n))$$

Dynamic Bayesian Networks

Dynamic Bayesian networks model probabilistic independencies over time. With certain assumptions, these can be seen as causal dependencies.



Data

Luminescence data was collected from experiments on *Arabidopsis Thaliana*. Data was collected on ten different genes. Each experiment had two phases – Entrainment and Constant light. Each experiments had different conditions that varied the light ratio in the Entrainment phase. Samples were taken every 1.5 hours for six days leading to 96 samples

Method

Five genes were selected, as these had the most biological information available as to their function

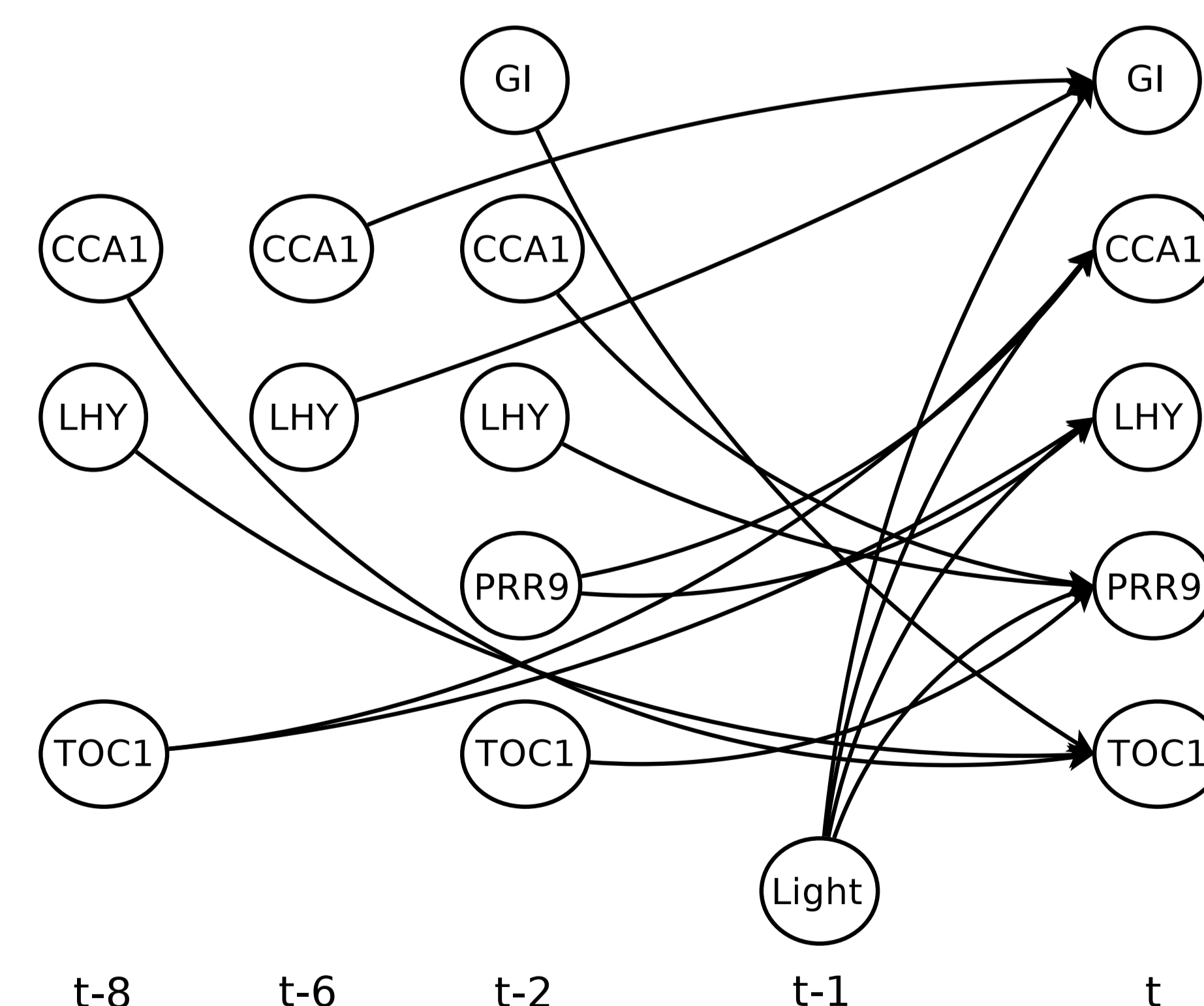
To fit Bayesian networks to the data, a score-and-search method was used with a meta-heuristic search criterion. The BDeu scoring function was used with a range of values for N' . This gives the relative posterior probability of the graph given the data, with Dirichlet priors and a uniform joint distribution assumption.

$$P(G, D) = P(D|G)P(G)$$

$$P(D|G) = \prod_{i=1}^n \prod_{j=1}^{q_i} \frac{\Gamma(N'_{ij})}{\Gamma(N'_{ij} + N_{ij})} \cdot \prod_{k=1}^{r_i} \frac{\Gamma(N'_{ijk} + N_{ijk})}{\Gamma(N'_{ijk})}$$

Evaluation

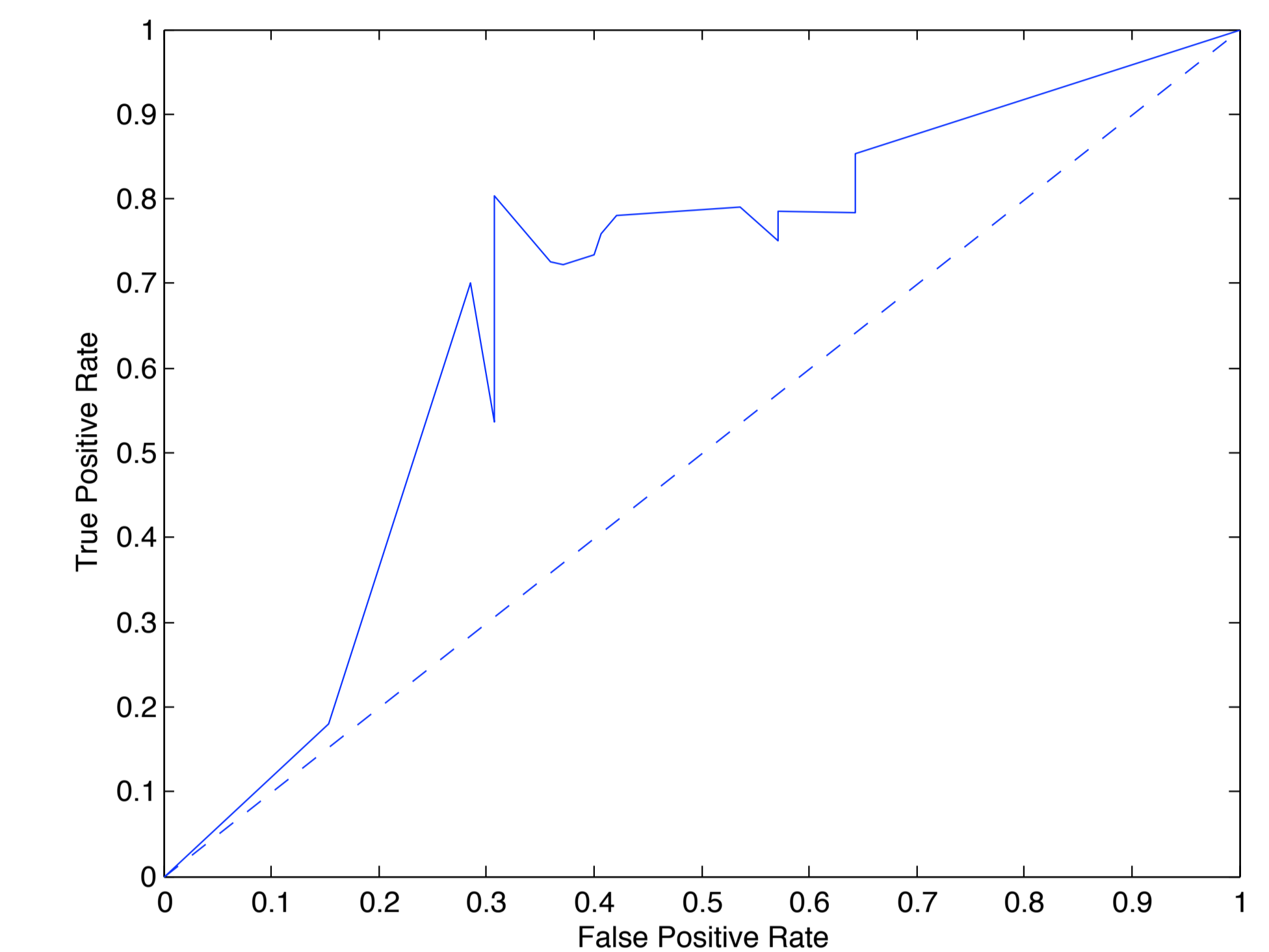
A dynamic Bayesian network was constructed based on hypothesised biological information. This network was used to compare against the learned networks.



Based on missing and present arcs, a true-positive rate and false-positive rate could be found. It was then possible to plots these rates on a ROC curve.

Results

Below are some sample results for one of the experiments.



It was also possible to average the DAGs learned over all the experiments and values of N' . This gave the following highly supported links.

Connection	Lag (hrs)
LHY → GI	9
CCA1 → GI	7.5
TOC1 → CCA1	12
TOC1 → LHY	12
CCA1 → PRR9	3
LHY → PRR9	3
CCA1 → TOC1	12
GI → TOC1	3

Connection	Lag (hrs)
PRR9 → LHY	12 (anti-phase)
LHY → CCA1	1.5 (similar)
GI → LHY	3 (spurious)

Acknowledgements

This work was supported by various grants from the BBSRC and EPSRC.