# Chapter 5
# Inferring Genetic Regulatory Interactions with Bayesian Logic-Based Model

**Svetlana Bulashevska**
*German Cancer Research Centre (DKFZ), Germany*

## ABSTRACT

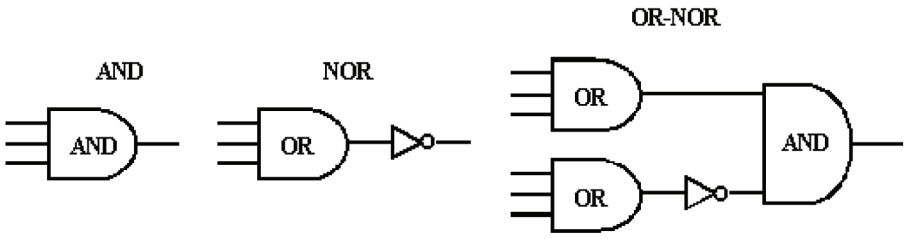
*This chapter describes the model of genetic regulatory interactions. The model has a Boolean logic semantics representing the cooperative influence of regulators (activators and inhibitors) on the expression of a gene. The model is a probabilistic one, hence allowing for the statistical learning to infer the genetic interactions from microarray gene expression data. Bayesian approach to model inference is employed enabling flexible definitions of a priori probability distributions of the model parameters. Markov Chain Monte Carlo (MCMC) simulation technique Gibbs sampling is used to facilitate Bayesian inference. The problem of identifying actual regulators of a gene from a high number of potential regulators is considered as a Bayesian variable selection task. Strategies for the definition of parameters reducing the parameter space and efficient MCMC sampling methods are the matter of the current research.*

## INTRODUCTION

The advent of microarray technology facilitated monitoring of gene expression and posed the problem of reconstructing genetic regulatory relations from data. A concept of *gene regulatory network* evolved, as a graphical representation of interactions between genes. This is a simplification of the underlying molecular biological regulatory mechanism, since the expression levels of some genes affect the expression of other genes indirectly, via the synthesis of proteins, protein complex formation, DNA binding etc. Mathematical models of genetic regulatory networks define features of the regulation by means of mathematical functions and propose algorithms in order to infer network models (i.e. connectivity, parameters etc.) from experimental data.

DOI: 10.4018/978-1-60566-685-3.ch005

*Figure 1. Examples of genetic regulatory functions presented as logic gates*




The attempt to model genetic regulation was pioneered long before the appearance of high-throughput molecular genetics methods (Kauffman 1969, 1996). It was stated that the regulatory interactions between genes can be presented as logic gates as exemplified in Figure 1, and the Boolean network model was proposed. In the Boolean network, discrete states of genes (the active and the not active) are admitted, and the state of each gene is functionally determined by the states of some other genes using the rules of logics. Continuous gene expression measurements must be discretized before they can be used for Boolean network modeling.

The fundamental idea behind the Boolean network is that the gene regulation is executed by transcription factors transcribed from a number of genes, which cooperatively bind to the binding sites of a target gene. This constitutes a so called *cis-regulatory element*, the working principles of which can be described by means of logics. Some genes are activated by one of several different possible transcription factors ("OR" logic). Other genes require that two or more transcription factors must all be bound for the activation ("AND" logic). The activation of some genes may be inhibited by one of a few possible repressor proteins ("NOT OR" logic, in our notation "NOR"). Further on, in case of "OR-NOR" logic, a gene is regulated by a set of possible activators and a set of possible inhibitors. The gene is transcribed if and only if one of its possible activators is active and it is not repressed by one of its possible repressors. An algorithm REVEAL was developed to reverse-engineer Boolean logic relations from expression data, based on mutual information between input and output states (Somogyi and Sniegosky, 1996; Liang et al., 1998). The major limitation of the Boolean network model was its inherent determinism, which contradicts with the stochastic nature of the underlying process of gene regulation and limits the reliability of relations inferred from real data.

Later on, extensions of Boolean Networks were suggested to make them robust against noise. In the *noisy Boolean networks* of Akutsu (2000), a certain probability is defined, with which a number of input/output patterns will not be discarded by an inference algorithm, even if a Boolean function is not satisfied. In the *Probabilistic Boolean Networks* (Shmulevich et al. 2002), more than one Boolean function are defined for each gene, and the particular function for calculating the state of the gene is selected with a certain probability.

Friedman et al. (2000) were the first to employ *probabilistic graphical models*, particularly *Bayesian networks*, to model genetic regulatory network. Probabilistic (statistical) modeling uses probability distributions to describe the states of the modeling variables and their dependencies. Probabilistic graphical models (Jordan, 2004) are graphs in which nodes represent random variables, and the missing edges between the nodes represent conditional independencies among the variables. In this way, the *joint probability distribution* of the variables is represented in a compact form. This reduces the number of parameters needed to describe the whole probabilistic model and sets a basis for statistical inference. Bayesian