

# Objectives of PNNL's Network Inference Testbed Project

The Network Inference Testbed (NIT) is being created as an interactive environment for the evaluation of algorithms used in the reconstruction of the structure of regulatory networks. Funding: the PNNL Biomolecular Systems Initiative (DOE).

The NIT

- compares and trains network inference methods on artificial networks and simulated gene expression perturbation data.
- allows the analysis of experimental high-throughput expression data using the suite of (trained) inference algorithms.
- will be of use to both software developers wishing to compare, refine, or combine inference techniques, and to bioinformaticians analyzing experimental data.

# Background (I)

Defining genetic regulatory networks is an important component of DOE's Genomes:GTL program. As stated by Frazier et al.: "An organism's genome also provides the basis for complex regulatory networks that control and coordinate the assembly and operation of these molecular machines. A second goal of GTL is to deconvolute and model these regulatory networks." [U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, 5]

## Background (2)

The structure of a genetic regulatory network can be described as a “wiring diagram,” a directed graph whose edges show regulatory influences. Such a diagram describes all the direct and indirect influences on the expression of a gene, and shows what (the product of) a gene can affect.

Ideally, one would like to have not just such a diagram, but also the set of equations governing the behavior of such a network over time. This would allow predictions to be made of the precise temporal behavior of all the (dependent) variables in the system. However, the diagram itself is an extremely important starting point for network analysis. It constrains the possibilities and shows what can affect different aspects of the system. It is the blueprint - the starting point for later investigations for dynamic behavior.

## Background (3)

This project assumes that NIT platform users seek to directly infer genetic regulatory networks from high-throughput microarray mRNA expression data, and protein regulatory networks from high-throughput protein expression data .

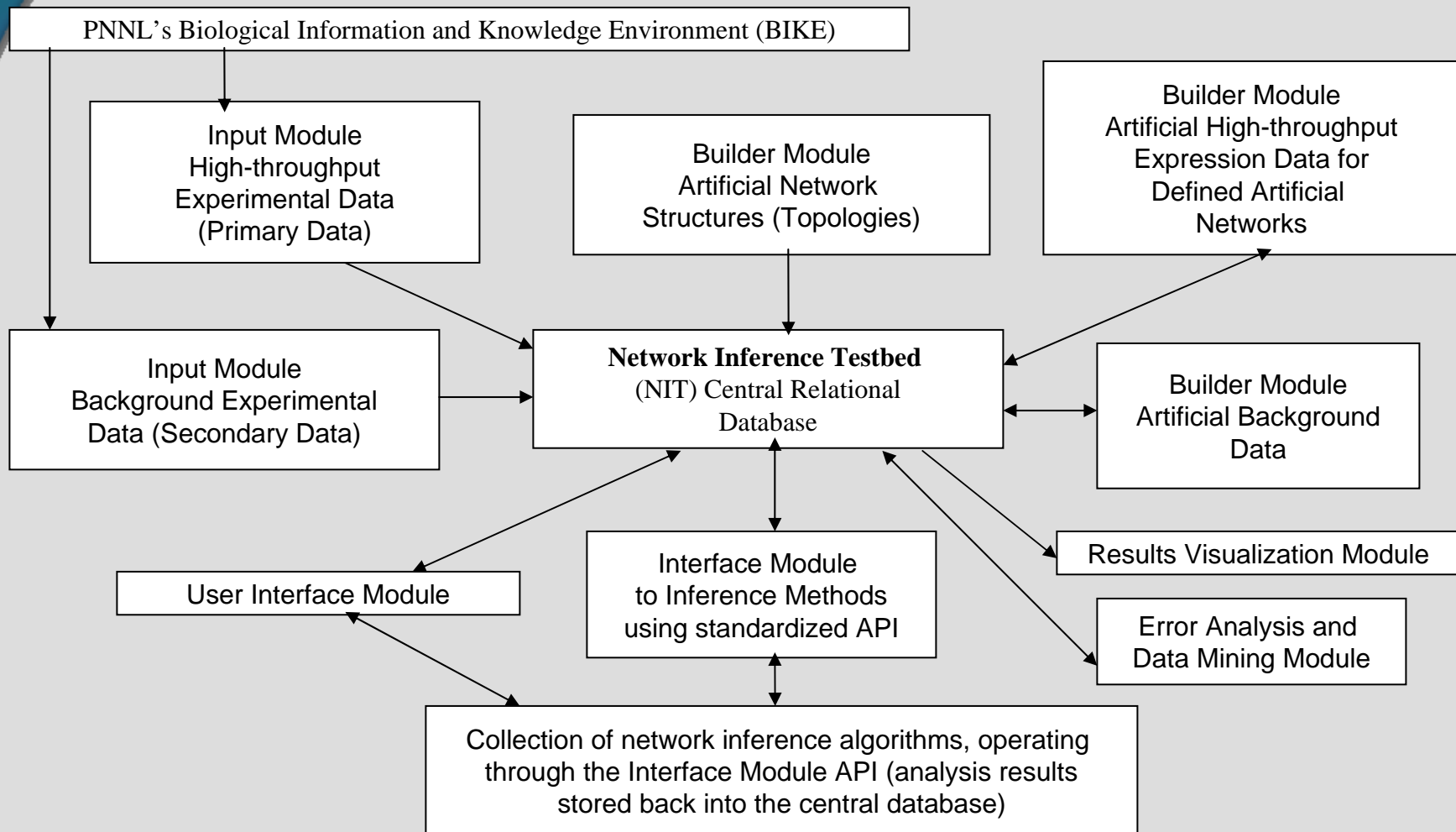
Briefly, methods using high-throughput data rely on searching for patterns of partial correlation or conditional probabilities that indicate causal influence. Such patterns of partial correlations found in the high-throughput data, possibly combined with other supplemental data on the network, will be the basis upon which the algorithms in the NIT's toolkit infer regulatory networks.

## **NIT architecture & artificial network use**

The NIT consists of a suite of independent software modules (Java) operating on a centralized relational database, controlled through a web interface.

It is being designed to allow direct comparisons of different inference algorithms on dynamically generated networks of different topologies. Also, the use of synthetic networks with corresponding artificial perturbation and time series expression measurement sets will allow machine learning inference techniques to undergo supervised training on known networks.

# NIT block diagram



# Summary - NIT capabilities

Thus, the testbed will allow

- ▶ Direct comparison of network inference methods on common data sets.
- ▶ Artificial data sets (topologies, perturbations, node input functions) that can be dynamically altered and stored.
- ▶ Inference results that can be stored, further analyzed, and visually displayed.
- ▶ Dynamic, step-wise refinement of inference methods, based on results.
- ▶ Well-defined addition of new inference algorithms through an API.
- ▶ Supervised or unsupervised training of inference methods, with supervised inference results scored against the known network topologies.
- ▶ Analysis of experimental data within the same framework.

Point of contact: [ronald.taylor@pnl.gov](mailto:ronald.taylor@pnl.gov)