INTRODUCTION
Chondrogenesis is a complex developmental process in cytoskeleton involving many transcription factors. Mathematical and computational models are useful to evaluate the large amount of microarray-derived expression data and predict the critical transcription-factor binding motifs. Using the temporal mRNA expression data and the regulatory DNA sequences from the human genome, we previously formulated a quantitative model to predict a set of critical transcription-factor binding motifs in a frame of combinatorial optimization problems [1]. In order to solve the problem, the efficient computational algorithm should be employed. Selecting 10 transcription-factor binding motifs from a random pool of DNA fragments of 5 bp in length, for instance, requires evaluation of $3.1 \times 10^{9}$ cases $(31,512)$ without considering DNA polarity.

In order to select the critical set of binding motifs, a particle swarm optimization procedure was applied. Swarm intelligence is an artificial intelligence technique that mimics a collective behavior in decentralized systems [2]. Such systems are made up by a population of individuals like ants and bees that interact locally with one another and with their environment. Using the mRNA expression data published by Sekiya et al. [3], the time dependent critical set of transcription-factor binding motifs in chondrogenesis was predicted using 200 artificial bees. Each bee represented a set of binding motifs, and bees flew in a solution space to minimize the model error through their communications.

METHODS
In the mathematical formulation, mRNA expression levels were modeled by the distributions of transcription-factor binding motifs in the 5'-end flanking DNA regions [4]. Five-hundred twelve motif candidates, 5-bp DNA fragments such as 5'-AAAAA-3', 5'-AAAAA-3', etc., were considered in the model. In order to avoid overfitting or underfitting the parameters, Akaike information criterion was used to estimate the number of binding motifs in the model. Here, the most important 10 binding motifs were selected [1].

In the procedure, each particle was treated as a single artificial bee in a swarm flying in the solution domain. It moved towards the best personal location and the best global location in steps 1 to 4.

1. A position of each particle is defined in a 512-dim solution space:
   \[ P = [p_1, p_2, \ldots, p_{512}] \] (1)
   where $p_i = 1$ (inclusion of the $i$th motif), and $p_i = 0$ (exclusion of the $i$th motif).

2. The model error, difference between the observed and predicted expression levels, was calculated for each particle.

3. The velocity of the particles was updated to reduce the model error:
   \[ V(t) = V(t-1) + \alpha \left[ \epsilon \left( p_{gbest} \right) - R \left( 1 - 1 \right) \right] + \epsilon \left[ p_{pbest} \left( (t-1) - p(t-1) \right) \right] \] (2)
   where $V$= velocity of the $i$th particle; $P_{gbest} =$ global optimal solution so far achieved; $P_{pbest} =$ local optimal solution so far achieved; $P =$ current solution of the $i$th particle; $\epsilon =$ velocity correction factor; $\epsilon =$ global contribution ratio; $\epsilon =$ two random numbers.

4. The particle was moved to a new position:
   \[ p(t+1) = p(t) + V(t) \] (3)

Steps 2-4 were iterated until the particles would converge to a global optimal position representing the 10 best binding motifs, or no improvement was achieved in a certain number of iterations.

RESULTS
The particle swarm optimization procedure predicted 10 binding motifs such as AACAT, AACTC, AGGCA, AGAGG, AGGGG, CATTC, CGTAC, CTCCC, CTGAC, and GGTAA. Using these 10 binding motifs, the predicted expression pattern for the 55 genes involved in chondrogenesis is illustrated (Fig. 1). The predicted regulatory network for type II collagen gene is shown including a linkage to the known binding motifs (Fig. 2).

DISCUSSION
The particle swarm optimization selected 10 binding motifs that shared part of existing motifs such as Ap1, Ap2, Lef1, Yy1, NF2B, and Sox9. Some of these motifs are known to be involved in chondrogenesis, and some others are novel. These novel motifs would generate testable hypotheses. Experimental evaluation such as a promoter competition assay [5], an electrophoretic mobility shift assay, and a gene reporter assay is required to validate the predicted transcription-factor binding motifs.

ACKNOWLEDGEMENTS
The authors thank I Sekiya and DJ Prockop for the microarray data. This study was supported by NIH R21RR17012.

REFERENCES