

R&D field: Life science

Validation of statistical significance for development of individualized medicine

- Program name: MLTest
- Developer
 - Naoyuki Kamatani, Prof. of Tokyo Woman's Medical Univ.
- Abstract
 - Calculation of generation probability of each pattern that distributes haplotypes to two ensembles based on haplotype frequencies of the ensembles.
 - Exact calculation of the type-I error rate.
- Algorithm
 - Calculation of generation probability of each pattern based on probability theory.
 - C and MPI.
 - Parallel computation of generation probability of each distribution pattern on each node.
- Current computation size
 - 5 haplotypes.
 - 50-member ensembles.
 - 1-day computation with Xeon 3.0GHz × 128 CPUs.
- Future computation size in 2010
 - Computation with the size of actual research for searching disease-related genes.
 - 10²-10³-member ensembles.

$$P = \left[\sum_{k_1=0}^{2n} \sum_{k_2=0}^{2n-k_1} \dots \sum_{k_k=0}^{2n-k_1-k_2-\dots-k_{k-1}} \binom{2n}{k_1} \binom{2n-k_1}{k_2} \dots \binom{2n-k_1-k_2-\dots-k_{k-1}}{k_k} \right] \\ \times h_1^{k_1} h_2^{k_2} \dots h_{k-1}^{k_{k-1}} h_k^{k_k} \\ \times \left[\sum_{k_1=0}^{2m} \sum_{k_2=0}^{2m-k_1} \dots \sum_{k_k=0}^{2m-k_1-k_2-\dots-k_{k-1}} \binom{2m}{k_1} \binom{2m-k_1}{k_2} \dots \binom{2m-k_1-k_2-\dots-k_{k-1}}{k_k} \right] \\ \times h_1^{k_1} h_2^{k_2} \dots h_{k-1}^{k_{k-1}} h_k^{k_k} \\ \times f(h_1, h_2, \dots, h_k)$$

The exact type-I error rate

- Expected results
 - Improvement of accuracy and acceleration of searching disease-related genes based on the exact calculation on the type-I error rate.
 - Realization of individualized medicine that represents anticipated side effects as probabilities based on individual gene types.
- Reference
 - <http://www.riken.jp/r-world/research/lab/idenn/statis/index.html>