

## **Artificial neural network used for the detection of mutations in DNA sequence raw data.**

Tobias Söderman and Lennart Björkesten,  
Amersham Pharmacia Biotech,  
SE-75184 Uppsala, Sweden.

*tobias.soderman@eu.apbiotech.com*  
*lennart.bjorkesten@eu.apbiotech.com*

Mutation detection has important applications in many different fields ranging from early drug discovery to clinical diagnostics. Even though there are many methods available there are few being as general in applicability and defining as accurately the nature of the change as direct DNA sequencing [1].

Depending on the origin of the mutation, one must in the general case expect any mixture between a mutated DNA component and the wild-type component in the sample. Inherited mutations typically give rise to a heterozygote 50-50 mixture while induced mutations e. g. in tumour tissue may give rise to any mixture depending on the heterogeneous composition [2].

Base-calling strategies are traditionally designed and optimised for genomics oriented applications. The main focus has been on read-length and on the accurate assignment of clean bases [3]. Less effort has been put into the analysis of heterozygote situations and general mixtures.

Our ANN based mutation detection algorithm is applied as a second pass on data processed by ordinary base-calling software. Every single base position is reconsidered as potentially hiding a point mutation. A number of descriptive features derived from the raw data traces, from the actual sample as well as from a reference sample, are fed into an artificial neural network, which is trained to produce mutation assignments.

The performance of the ANN based mutation detection algorithm will be discussed in terms of sensitivity and specificity and it will be pointed out that pre-processing of the descriptive features is essential in order to keep the required training at a reasonable level.

### **References**

- [1] Grompe, M., Nat. Genet. 5, 111-117 (1993).
- [2] Hederum, A., Pontén, F., Ren, Z., Lundeberg, J., Pontén, J. and Uhlén, M., BiTechniques 17, 118-129 (1994).
- [3] Tibbetts, C., Bowling, J. M. and Golden, J. B. III. (1994). In Adams, M. D. et al., eds., Automated DNA Sequencing and Analysis, 219-230. San Diego, CA: Academic Press.