



## Plant & Animal Genomes XVI Conference

January 12-16, 2008  
Town & Country Convention Center  
San Diego, CA

---

P873 : Software

---

### A Perl Based Data Mining Workflow For Animal Breeding - From Phenotype To SNP

[Hendrik-Jan Megens](#)<sup>1</sup>, [Richard P.M.A. Crooijmans](#)<sup>1</sup>, [Barbara Harlizius](#)<sup>3</sup>, [Sonja Kollers](#)<sup>3</sup>, [Erik Mullaart](#)<sup>4</sup>, [Annemieke Jungerius](#)<sup>2</sup>, [Hindrik H.D. Kerstens](#)<sup>1</sup>, [Martien A.M. Groenen](#)<sup>1</sup>

<sup>1</sup> Animal Breeding and Genomics Centre, Wageningen University, Marijkeweg 40, 6709 PG, Wageningen, The Netherlands

<sup>2</sup> Hendrix Genetics P.O. Box 114, 5830AC Boxmeer, The Netherlands

<sup>3</sup> Institute for Pig Genetics, P.O. Box 43, 6640 AA Beuningen, The Netherlands

<sup>4</sup> CRV, P.O. Box 5073, 6802 EB, Arnhem, The Netherlands

With vast and rapidly increasing resources of genomic data available, there is a challenge for breeders to utilize this information in their breeding practices. While tools to combine multiple databases to aid in selecting genes and SNPs for medical studies have become publicly available, none exist that can comprehensively and species specifically do this for livestock animals. A consortium of Wageningen University and three major Dutch animal breeding companies aims to develop tools for mining genomic data specifically for breeding purposes. This is done by a Perl-based pipeline integrating genomic data, both from livestock and model species, from multiple databases, such as Ensembl and NCBI. The pipeline has a modular design which facilitates its use as modular tool development platform for our genomics research. A CGI-based interface provides user-friendly access to the tool (log-in is currently required). Based on PubMed, GO, gene-network information and QTL data, multiple candidate genes are automatically selected. These candidate genes are subsequently searched for putative functional information. To this end, several databases (e.g. MiRbase) and programs (e.g. clover, footprinter, miRanda) are implemented to find putative functional sites in or near the gene such as miRNA and transcription factor binding sites. Finally, public SNP databases are searched for variation at these functional sites. The main objective of the tool is the development of large SNP sets consisting of putative functional SNPs, thus facilitating a high throughput implementation of the candidate gene approach.