

Antigen selection for antibody-based proteomics

Lisa Berglund, Erik Björling, Linn Fagerberg, Kalle Jonasson, Åsa Sivertsson, Johan Rockberg, Cristina Al-Khalili Szigyarto and Mathias Uhlén

School of Biotechnology, Royal Institute of Technology, Stockholm, Sweden

Abstract

The completion of the human genome sequence has provided new opportunities for systematic exploration of the human proteome. One approach, taken by the Swedish Human Protein Atlas (HPA) program, is to generate protein-specific antibodies towards one representative protein from each human gene and to use these antibodies to analyze tissue- and cell protein profiles. One of the challenges in antibody-based proteomics is the high-throughput selection of suitable antigens. In HPA, antigens coined Protein Epitope Signature Tags (PrESTs) are 25-150 amino acids fragments of the target protein. An interactive visualization tool, PRESTIGE, has been developed in-house to facilitate *in silico* PrEST design by displaying protein feature data from both public and internal sources. The criteria set for PrEST selection, including low sequence similarity of the PrEST to other proteins, absence of transmembrane regions, and sufficient PrEST length, allows for selection of PrESTs on ~80% of the human proteome. To date, 27,000 PrESTs have been designed in HPA, covering more than half of the human genes.

Antigen selection

PrEST design in HPA is based on:

- low sequence identity of the selected region to proteins from other genes, for generation of specific antibodies;
- avoidance of signal sequence region (often cleaved from the mature protein) and transmembrane regions (not accessible for the antibodies *in vivo*);
- no NotI/Ascl restriction sites due to the use of these restriction sites in subsequent cloning of the PrEST into a vector for protein expression;
- PrEST length of 25 – 150 amino acids;
- preferably 4 PrESTs per protein, for validation purposes.

The PRESTIGE software facilitates selection of antigens using the above given criteria (Fig. 1), with data from the Ensembl database (www.ensembl.org).

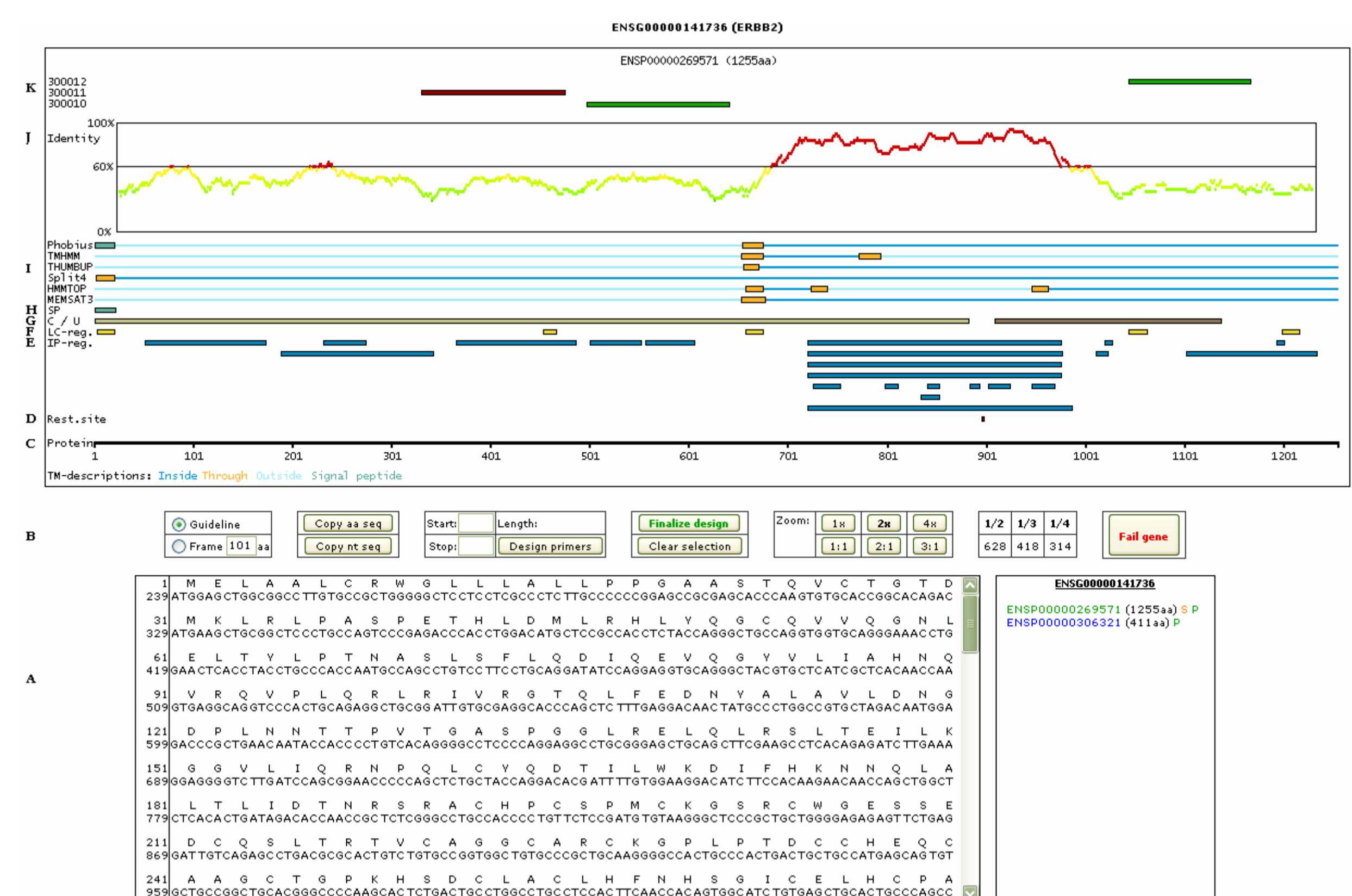


Figure 1. PRESTIGE – a tool for *in silico* selection of antigens. **A.** Amino acid sequence of the protein, and nucleotide sequence of the coding part of the processed transcript. **B.** Interactive features for sequence retrieval, position selection, primer design, zooming etc. **C.** Protein scale. **D.** Ascl/NotI restriction sites. **E.** InterPro regions. **F.** Low complexity regions. **G.** Common/unique regions between alternative splice variants. **H.** Signal peptide prediction. **I.** Topology prediction for membrane proteins, using different methods. **J.** Sequence identity to proteins from other genes, based on 51 amino acids sliding windows. **K.** Selected antigens.

Current status

15,000 human genes have been initiated in the HPA pipeline, corresponding to 27,000 selected antigens. The fraction of started genes over all human chromosomes is shown in Fig. 2.

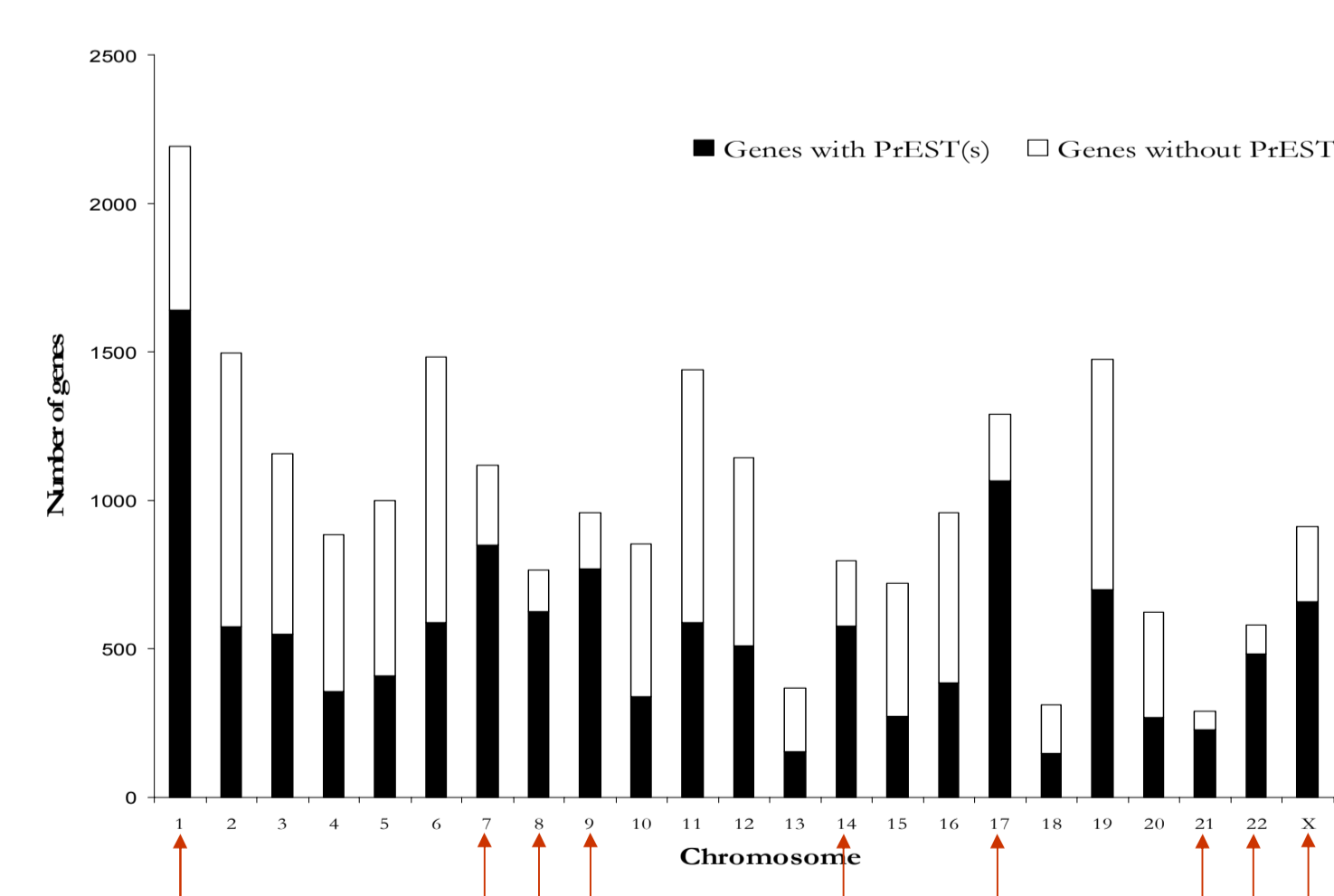


Figure 2. Fraction of human genes with PrEST(s). Genes on chromosomes marked with a red arrow have been subjected to systematic antigen selection.

Acknowledgements

This work was funded by the Knut and Alice Wallenberg Foundation.

