Molecular identification of genes and pathways involved in skeletogenesis by EST generation, full length cDNA isolation and expression profiling

Bernhard Zabel [1], **Erwin R. Schmidt** [2], **Thomas Hankeln** [2], **Andreas Winterpacht** [3]
[1] Univ.-Kinderklinik, Johannes-Gutenberg-Universität Mainz, 55101 Mainz [2] GENterprise – Gesellschaft für Genanalyse und Biotechnologie mbH, 55128 Mainz [3] Friedrich-Alexander-Universität Erlangen, Institut für Humangenetik, 91054 Erlangen

Introduction

The project was designed to identify and analyse genes and gene networks to better understand the complex processes responsible for both, normal and pathological skeletal growth, development and homeostasis. The work has its main focus on human disorders and will take advantage of the mouse as a model system for functional studies. Our results should provide both, an essential tool to investigate the molecular basis of numerous inherited disorders of the growing skeleton, as well as crucial information on mechanisms involved in cartilage/bone remodeling and degeneration as observed in osteoarthritis, osteoarthrosis and osteoporosis. The research in this field was hampered by the fact that cartilage - and especially human cartilage - has restricted accessibility and is a scarce source of RNA. Consequently, this tissue has not been included in most EST projects, and has not been investigated systematically in patient tissues by expression profiling. Our work is founded on the only human fetal cartilage cDNA library available at that point (provided to us by Dr. B. Lee, Baylor College of Medicine, Houston). The implementation of the proposed strategy was successful and the initial goals have been reached. New scientific aspects (see below) resulted in some changes of time schedule and in the final program of the project.

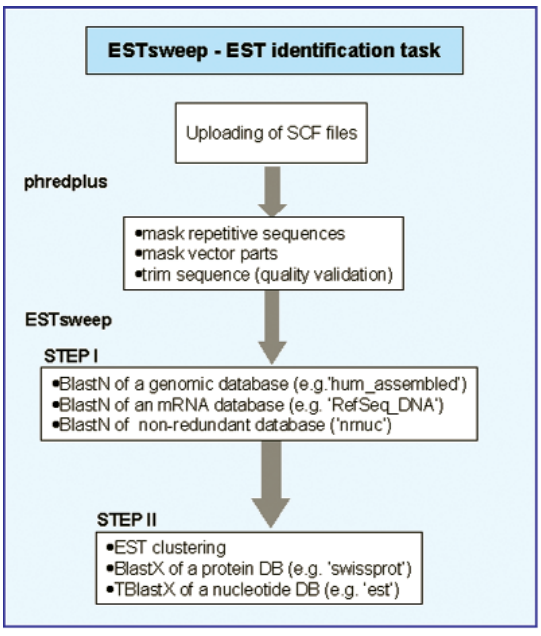

Fig 1: "ESTsweep"

EST-Sequencing and bioinformatical evaluation

During the EST sequencing phase, we have generated a total of 5.000 sequences from the human fetal cartilage cDNA library. The average read length of the ESTs was 500 bp. To obtain this number of quality-controlled ESTs, we had to prepare more than 15.000 templates due to the high number of very short inserts within the library.
The sequence reads have been up-loaded on a FTP server, available to all members of our consortium. In order to facilitate the bioinformatical evaluation of the EST reads, we devised a procedure, which includes the necessary steps of trimming sequences and database searching. It was set up and optimized in close collaboration with Dr. Hotz-Wagenblatt and Dr. Glatting from the HUSAR (Heidelberg Unix Sequence Analysis Resources) bioinformatics team at the DKFZ in Heidelberg. The EST evaluation routine (called "ESTsweep", see Fig.1) is available to the public through the HUSAR program package (http://genome.dkfz-heidelberg.de). (Fig.1)

So far, about 4.000 EST sequences have been run through ESTsweep. Most of the EST sequences (69,6%) show significant similarity to known genes of the human RefSeq collection (NM_#), and another 4,8% correspond to human model RNAs (XM_#). 2,6 % of the sequences are contaminated with bacterial or vector sequences. The remaining 23% (9% genomic, 14% anonymous EST match/no matches) have to be further analysed, because they do not show any similarity to defined mRNAs or proteins, but to genomic clones or anonymous ESTs. The current classification of the analysed EST-sequences is summarized in Fig.2.

The clones representing sequences of unknown coding potential are currently under detailed scrutiny by further bioinformatical tools, since they may contain candidates for novel transcripts.

Characterization of selected genes

A number of about 20 novel candidate genes are currently under intensive investigation by detailed bioinformatical and experimental methods. The latter includes the verification of tissue specificity and the reconstruction of possible transcripts by RT-PCR. In some cases, the subcellular localization of the putative protein has been determined and first expression studies in mouse or human have been performed by Northern blotting and mRNA in situ hybridization.

Candidate gene identification and characterization
Reaching a stage of the project when identification of candidate genes involved in cartilage and bone formation proves to be a key issue, we established a cooperation with the Clinical Research Unit "Osteogenic stem cell differentiation and therapy of bone loss" of Prof. Jakob in Würzburg. They focus on the highly innovative field of human mesenchymal stem cells (hMSC) differentiation into cells of the osteogenic and chondrogenic pathways. There are mutual benefits from the joint effort with this research unit. We would be able to use their system to identify cartilage and bone specific transcripts out of our EST pool and they would in return obtain a systematic molecular characterization of the hMSC differentiation process, with series of new skeletal candidate genes, generated by the EST project. The collaboration will have additional advantages by using the hMSC cultures as reliable human system for further gene function studies and for the generation of bone and cartilage related full length cDNA libraries.

Working program based on incorporation of the hMSC system into the project
In order to characterize our EST dataset and to select candidate genes for further analysis, we will pursue an updated working program. It will include: (1) The characterization of the pattern of EST expression in a reproducible model of target tissues, exploiting the in vitro differentiation of hMSC into mature cells of the target tissues (e.g. osteoblasts, chondrocytes and adipocytes) with verification of differential expression by RT-PCR and Northern analysis. (2) Selection of candidates according to their expression pattern and to in-silico data. (3) Establishment of cDNA-libraries from undifferentiated and differentiated hMSC-cultures to gain access to full-length transcripts. (4) Cloning und sequencing of selected candidates as well as recombinant expression und antibody-production, biological and functional characterization. (Fig.3)

A schematic overview of this project strategy is given in Fig. 3. Original project features like EST-sequencing, characterization of human (and corresponding mouse) genes are complemented by the hMSC system as basis for EST screening procedures, full length cDNA generation and microarray expression studies. Strong skeletal candidate genes may proceed to further levels of gene function analysis including the use of murine model systems, mutation screening in patients and recombinant gene expression studies.

1 deCrombrugghe B *et al.*: **Regulatory mechanisms in the pathways of cartilage and bone formation.** Curr Opin Cell Biol 2001, 13: 721-7,

2 Hall C *et al.*: **International Nosology and Classification of Constitutional Disorders of Bone.** Am J Med Genet 2002 (in press)

3 Stokes DG *et al.*: **Assessment of the gene expression profile of differentiated and dedifferentiated human fetal chondrocytes by microarray analysis.** Arthritis Rheum 2002, 46: 404-19

4 Deans RJ, Moseley AB: **Mesenchymal stem cells: Biology and potential clinical uses.** Exp Hematol 2000, 28: 875-84

5 Nöth U *et al.*: **Trabecular bone derived mesenchymal stem cells.** J Orthop Res 2002 (in press)

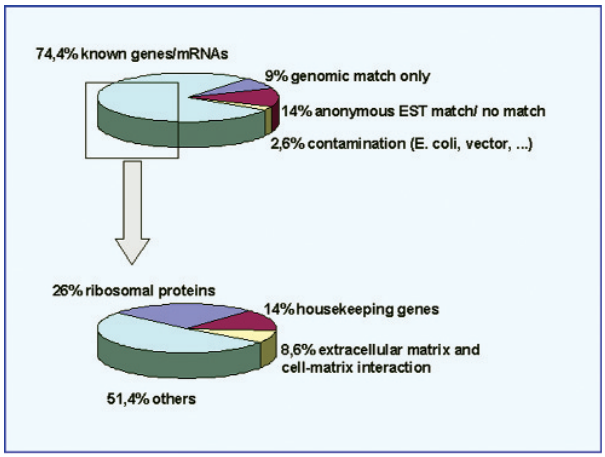

Fig 2: **Bioinformatic analysis of 4.000 EST clones**

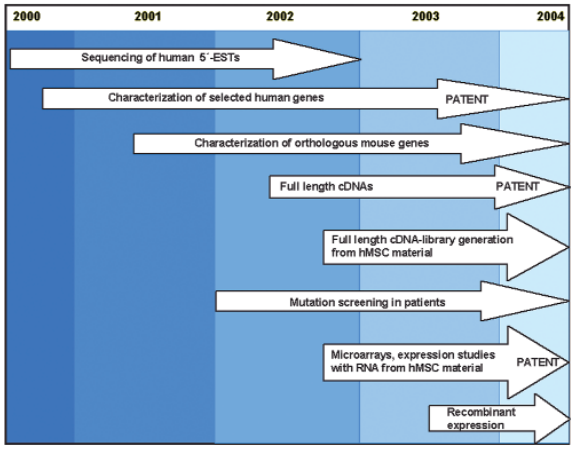

Fig 3: **Extended and updated working plan**