This is the author's version of a work that was submitted/accepted for publication in the following source:

Friis, Thor Einar, Stephenson, Sally-Anne, Xiao, Yin, Whitehead, Jon, &
Hutmacher, Dietmar W.
(2014)
A polymerase chain reaction-based method for isolating clones from a complimentary DNA library in sheep.
Tissue Engineering Part C: Methods, *20*(10), pp. 780-789.

A Polymerase Chain Reaction-Based Method for Isolating Clones from a Complimentary DNA Library in Sheep

AU1 ▶

AU2 ▶

Thor Einar Friis, PhD,[1] Sally Stephenson, PhD,[2] Yin Xiao, PhD,[3]
Jon Whitehead, PhD,[4] and Dietmar W. Hutmacher, PhD[1]

The sheep (*Ovis aries*) is favored by many musculoskeletal tissue engineering groups as a large animal model because of its docile temperament and ease of husbandry. The size and weight of sheep are comparable to humans, which allows for the use of implants and fixation devices used in human clinical practice. The construction of a complimentary DNA (cDNA) library can capture the expression of genes in both a tissue- and time-specific manner. cDNA libraries have been a consistent source of gene discovery ever since the technology became commonplace more than three decades ago. Here, we describe the construction of a cDNA library using cells derived from sheep bones based on the pBluescript cDNA kit. Thirty clones were picked at random and sequenced. This led to the identification of a novel gene, C12orf29, which our initial experiments indicate is involved in skeletal biology. We also describe a polymerase chain reaction-based cDNA clone isolation method that allows the isolation of genes of interest from a cDNA library pool. The techniques outlined here can be applied in-house by smaller tissue engineering groups to generate tools for biomolecular research for large preclinical animal studies and highlights the power of standard cDNA library protocols to uncover novel genes.

Introduction

THE SHEEP (*Ovis aries*) has many features that makes it highly suited as a large animal model for orthopedic research. Sheep have a body weight that is similar to humans and the size of their long bones allows the use of implants and internal fixation plates used in human clinical practice. Compared with other large animal models, such as dogs, goats, pigs, and non-human primates, sheep have docile temperaments, are easy to house, and relatively inexpensive to acquire and maintain. For these reasons the sheep is increasingly becoming the animal model of choice for bone tissue engineering research (reviewed in detail by Refs.[1–4]). The first sheep genome draft was released in 2009,[5] however, there are still gaps in the annotated sequence information available in the sheep transcriptome, particularly when compared to other species, such as the cow, mouse, rat, or humans. Despite the great utility of sheep as an animal model, there is a lack of molecular tools that can be applied for musculoskeletal tissue engineering research such as sheep-specific antibodies or polymerase chain reaction (PCR) primers.

The construction of a complimentary DNA (cDNA) library reflects the gene expression from tissues or tissue-specific cells at a given time and is a valuable resource for biomolecular applications. The conversion of mRNA into cDNA is performed by a series of enzymatic reactions that involves the synthesis of first strand cDNA from the mRNA template with a reverse transcriptase followed by the conversion to a double-stranded cDNA and insertion into a cloning vector.[6] The construction of a cDNA library can be readily performed by a laboratory with the necessary molecular biology facilities but for a tissue engineering group the task can represent a technical challenge.

In broad terms there are two different approaches one can take when performing cDNA library experiments. There is the high throughput approach, typically in the form of expressed sequence tags (ESTs), which are short (100–800 bp), randomly selected single-pass sequence reads derived from cDNA libraries.[7] More recently, whole transcriptome sequencing, using next-generation sequencing (NGS) platforms such as Illumina, SOLiD, and 454 FLX, can sequence each transcript at a depth of 100–1000 reads, often as short as 50–500 bases. The NGS strategy heavily relies on advanced bioinformatics tools and computer resources to manage and analyze the terabytes of raw data often produced with these technologies.[8] For research groups without the infrastructure, resources and skills sets needed for the

[1]Regenerative Medicine Group, Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, Queensland, Australia.
[2]Cells and Tissue Domain and [3]Bone Group, Institute of Health and Biomedical Innovation, Queensland University of Technology, Kelvin Grove, Queensland, Australia.
[4]Metabolic Medicine Group, Mater Medical Research Institute, South Brisbane, Queensland, Australia.

NGS strategy, a conventional cDNA library strategy coupled with standard Sanger sequencing is still a valid means of adding to the body of genetic information for a given model species.

In this study, we describe the construction and characterization of a cDNA library using cells derived from sheep bones with an emphasis for tissue engineering groups that may not be specialized in molecular biology. Using this approach, we isolated the uncharacterized gene, *C12orf29*, and present our initial data that indicate it has a key role in skeletal biology. We also describe, in detail, the protocol used to isolate the full-length cDNA of ovine glyceraldehydes-3-phosphate dehydrogenase (GAPDH) using an inverse PCR method. This study demonstrates the utility of conventional cDNA library technologies to uncover novel genes that can greatly add to our basic knowledge of biology, the ground well from which biomedical sciences is ultimately informed.

Materials and Methods

Cell culture

Primary cell lines were established by explant cultures from sheep bones from mandible and tibia and bone marrow. The use of these samples had been approved by the Queensland University of Technology Animal Ethics committee (approval no. 700000915). The bone samples were processed by first removing the soft tissues; care was taken to remove all periosteal membranes. The bones were fragmented into smaller pieces, $\sim 15 \times 6 \times 6$ mm, using a bone rongeur. The fragments were washed thrice in sterile phosphate-buffered saline (PBS), transferred to a clean 50 mL tube, and washed in 30% ethanol by vigorous shaking for 3–4 min. The ethanol was removed and the fragments were washed thrice in sterile PBS before being incubated for 10 min in 5 mL of 0.25% trypsin (Life Technologies) at 37°C. The fragments were washed again in sterile PBS to remove any residual trypsin, then transferred to a 175 cm^2 (T175) tissue culture flask to which 25 mL of complete medium (Dulbecco's modified Eagle's medium [DMEM] containing 10% v/v fetal calf serum [BioWhittaker] 50 U/mL penicillin, 50 μg/mL streptomycin, and 0.5 μg/mL Fungizone [Life Technologies]) was added. Osteoblast cells typically emerged from the bone fragments within 7 days of being explanted (Fig. 1A). Once confluent, the cell numbers were expanded, then trypsinized, and stored in liquid nitrogen for later use. ◀F1

Bone marrow stromal cells were isolated from bone marrow from the iliac crest while the animals were under general anesthesia. Five milliliter of bone marrow was diluted in an equal volume of PBS in a 50 mL Falcon tube and 10 mL of Lymphoprep (Axis-Shield PoC AS) carefully
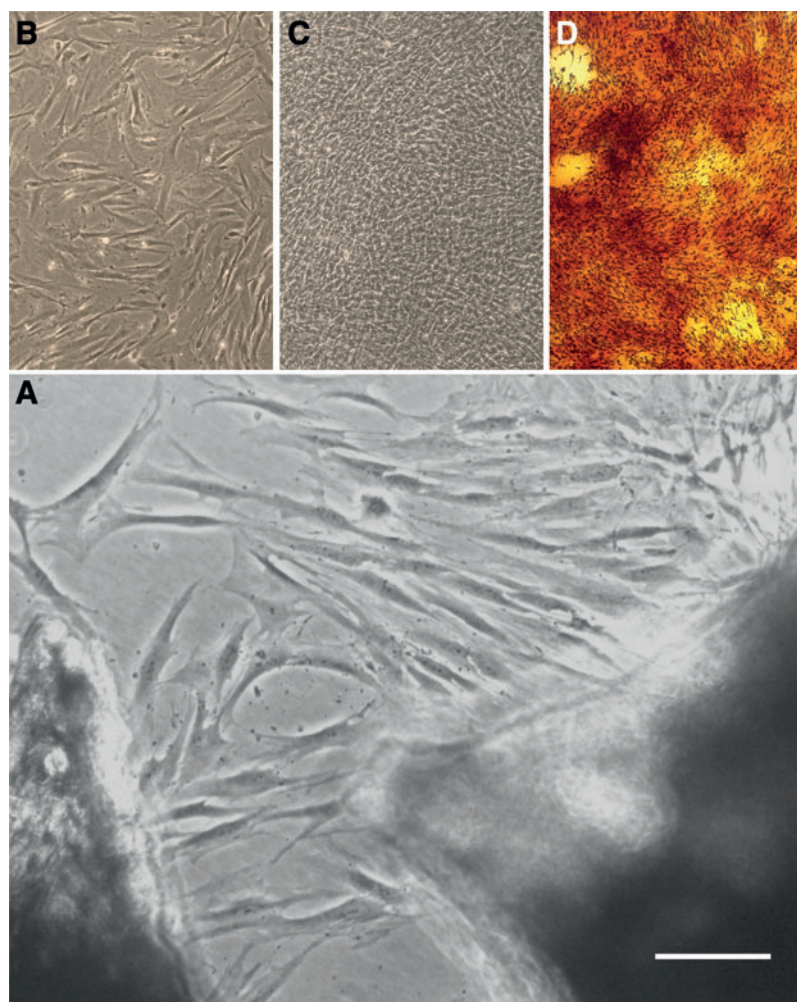

FIG. 1. Bone explant cell culture and mineralization induction. (**A**) Osteoblast cells would typically emerge from the sheep bone fragments after 7 days in culture. Scale bar = 300 μm. (**B**) Mineralization induction medium was applied to the p1 cells at ~70–80% confluence. (**C**) The cells responded by laying down a crisscross patterned extracellular matrix following mineralizing media treatment. (**D**) Following 3 weeks of growth in mineralization medium there was robust evidence of mineralization as seen by Alizarin red staining. Color images available online at www.liebertpub.com/tec

placed underneath the bone marrow with a serological pipette. The erythrocytes were pelleted by centrifugation at 800 g for 20 min at room temperature. The upper layer of the column ($\sim$10 mL) was removed and mixed with 30 mL of complete medium. Cells were plated out into T175 tissue culture flasks and unattached hematopoietic cells were removed by media change 24–48 h later. The cells were grown until confluency, typically 5×10^6 cells per flask, and then trypsinized and stored in liquid nitrogen for later use.

Mineralization induction

Three cell lines each from mandible, tibia, and BMSCs at passage 1 were plated out in T175 flasks and grown to 70–80% confluence (Fig. 1B) before being treated with mineralizing induction medium consisting of complete media containing 100 nM dexamethasone, 10 mM β-glycerophosphate, and 50 μg/mL L-ascorbic acid (Sigma-Aldrich)[9] for 3, 7, and 12 days (Fig. 1C). Mineralization controls were cultured in mineralization medium for 3 weeks and assessed by Alizarin red staining (Fig. 1D).[10]

RNA extraction and poly(A)$^+$ mRNA isolation

Total RNA was extracted at each time point using an MN NucleoSpin RNA L Midi Kit (Scientifix) and the poly(A)$^+$ mRNA fraction was enriched using an MN NucleoTrap poly(A)$^+$ RNA kit (Scientifix) following the manufacturer's instructions. The samples from the respective time points were pooled, then ethanol precipitated and resuspended into RNAse-free water, and resuspended to a final volume of 250 ng/μL.

cDNA library construction

cDNA was prepared from 5 μg of poly(A)$^+$ enriched RNA and was directionally ligated into the pBluescript II SK (+) XR vector, then chemically transformed into XL-10 Gold competent cells for propagation (Integrated Sciences, NSW, Australia) following the manufacturer's protocol.

Isolation of random clones from the library

Random colonies were picked from selective ampicillin agarose plates and propagated in Luria-Bertani (LB) medium with 100 μg/mL ampicillin. The plasmids were extracted using MN Nucleospin Miniprep kit (Scientifix) and analyzed by restriction enzymes to determine the presence of cDNA inserts. Positive clones were sequenced using M13 universal primer sites flanking the multiple cloning site. The samples were submitted to the Australian Genome Research Facility (AGRF) for sequencing.

Sequence assembly and analysis

The sequences were BLAST searched to identify homologous genes in human and cow. Formatting and identification of open-reading frames (ORFs) was performed using the Sequence Manipulation Suite (www.bioinformatics.org/sms2/). The sequences were assembled using Microsoft Word 2010 and prepared for submission to GenBank using the Sequin program, which was downloaded from www.ncbi.nlm.nih.gov/Sequin/

Isolation of specific clones using the MACH protocol

Validation of the library was done by isolating a full-length clone of *GAPDH*, and a partial clone of transforming growth factor beta 3 (*TGF-β3*; data not shown). These clones were isolated using a modification of the MACH cDNA clone isolation protocol.[11] This is a PCR-based protocol that makes use of two sets of abutting primers for inverse PCRs; the PCR products are subsequently separated by gel electrophoresis and annealed by base pair complementarities (Fig. 2A).

◀F2

Primer design

The mRNA sequences of human *GAPDH* (NM_002046) was BLAST searched against the sheep virtual genome on the International Sheep Genome Consortium web portal (https://isgcdata.agresearch.co.nz/) to identify a suitable stretch of sequence for primers (Fig. 2B). The MACH primers are designed to abut with no overlaps such that reverse primer 1 (R1) abuts forward primer 2 (F2) and reverse primer 3 (R3) abuts forward primer 4 (F4). The primers pairs (R1/F2 and R3/F4) can be adjacent to each other such that primers F2 and R3 are complementary as was the case with the *GAPDH* MACH primers but can also be offset such that there is no overlap between F2 and R3. The outside primers (OS5′ and OS3′) flank the area of the MACH primers (Fig. 2C) and were used to identify the relative abundance of the gene of interest (GOI). The primer sequences are shown in Table 1.

◀T1

PCR setup

All primers (Geneworks) were diluted to a working concentration of 10 μM and for each target gene 300 ng of cDNA library template was used in two separate reactions with primers R1/F2 and R3/F4. Each reaction was set up in a final volume of 50 μL using Phusion Hot Start II (Genesearch) following the manufacturer's instructions omitting the DNA polymerase. The reactions were preheated for 3 min at 98°C, after which 5 U (2.5 μL) of DNA polymerase was added per reaction. A touchdown PCR protocol was applied where the initial annealing temperature (Tm) was set 15°C above that of the lowest *T*m of the primer sets and lowered 1°C per cycle over 15 cycles.[12] The thermal cycling parameters are shown in Table 2.

◀T2

Dpn*I* digest

The PCR products were column purified using a MN NucleoSpin Extract II kit (Scientifix) and eluted in 25 μL final volume. The template cDNA, which is methylated, was digested with *Dpn*I restriction enzyme (Genesearch); the amplified PCR product, which is not methylated, is not digested by this restriction enzyme and therefore remains intact.

Gel separation and purification

The digested samples were loaded onto a 0.7% agarose gel and separated by electrophoresis. Equal sized DNA fragments were excised from the gels (Fig. 2A, step 3) and column purified and a 1 μL aliquot from each sample separated by gel electrophoresis for quality control.

FIG. 2. Polymerase chain reaction (PCR)-based method for isolation of specific clones from a complimentary DNA library pool. (A) Two PCRs were performed using two sets of inverse and abutting primers. The parent library was digested with *Dpn*I restriction enzyme after which the PCR products are separated on an agarose gel and bands of equal sizes (∼5 kb) cut out and purified. The purified PCR products were combined in a single tube and denatured and annealed. Only strands with complementary overhangs, that is, 1+4 or 2+3, formed nicked circles, when transformed into competent cells, which were repaired and replicated as circular plasmids. (B) The coding sequence of human glyceraldehydes-3-phosphate dehydrogenase (*GAPDH*) mRNA sequence was BLAST searched against the sheep genomic database to find an appropriate region in which to place the MACH primers. (C) The arrangement of the forward and reverse MACH primers is seen here. Two separate PCRs are set up with the inverse and abutting primer pairs R1/F2 and R3/F4. The outside (OS) primers are used to determine the abundance of the clone of interest. Color images available online at www.liebertpub.com/tec
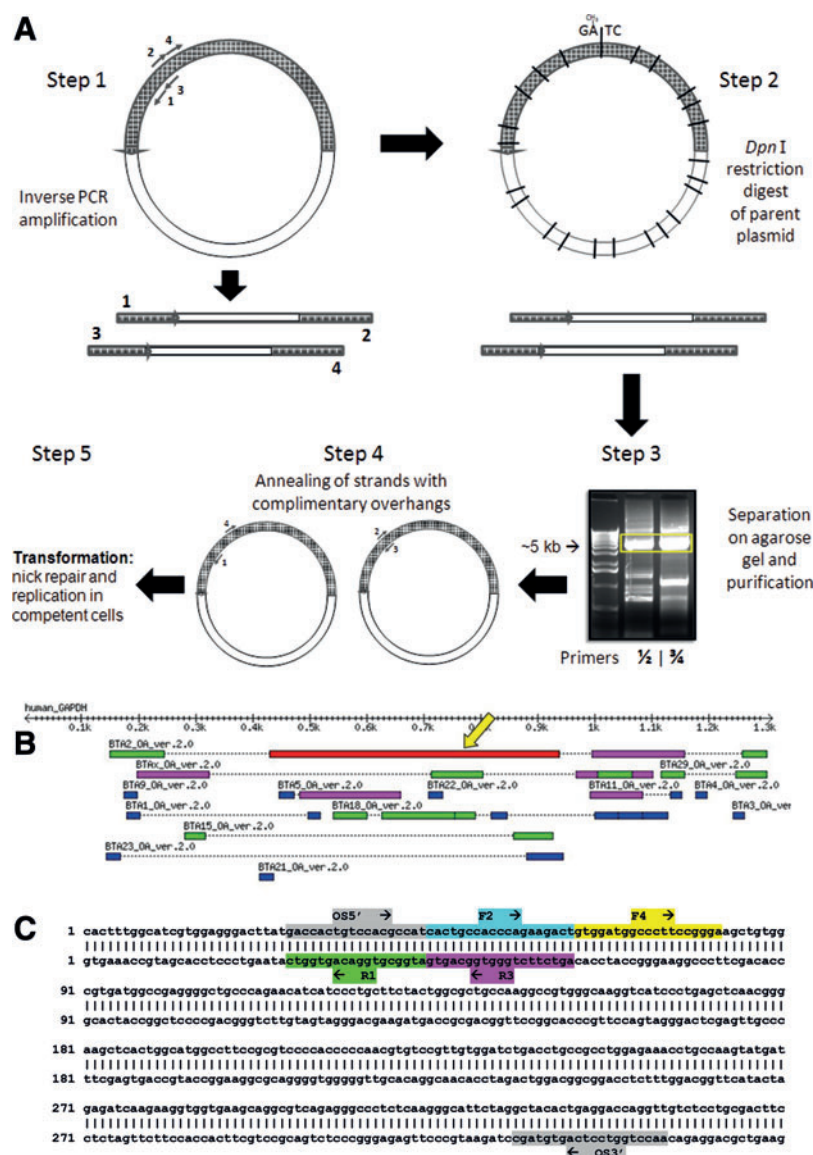
Annealing step

Equal volumes (6 μL) of R1/F2 and R3/F4 purified PCR product was combined with 1.5 μL 10×NEB 2 buffer and 1.5 μL of water. The fragments were denatured once at 95°C for 5 min, and then annealed over four cycles at 65°C for 2 min and 25°C for 15 min. The PCR products can only circularize by forming complementary overhangs between

R1+F4 and R3+F2 (Fig. 2A, step 4), which for the GAPDH primers was 21 bases long.

Transformation, plasmid extraction, and screening by analytical digest

The circularized plasmids were transformed into chemically competent α-Select Competent cells (Bioline) and the

TABLE 1. OLIGONUCLEOTIDE PRIMERS USED FOR MACH CLONAL ISOLATION PROTOCOL

Primer	Sequence	Length	GC content (%)	T_M (°C)
GAPDH OS5′	5′-GACCACTGTCCACGCCAT	18	61.10	57.8
GAPDH OS3′	5′-AACCTGGTCCTCAGTGTAGC	20	55.00	56.5
GAPDH R1	5′-ATGGCGTGGACAGTGGTC	18	61.10	57.8
GAPDH F2	5′-CACTGCCACCCAGAAGACT	19	57.90	57.1
GAPDH R3	5′-AGTCTTCTGGGTGGCAGTG	19	57.90	57.1
GAPDH F4	5′-GTGGATGGCCCTTCCGGGA	19	68.40	62.5

GAPDH, glyceraldehydes-3-phosphate dehydrogenase.

TABLE 2. TOUCHDOWN POLYMERASE CHAIN REACTION
TEMPERATURE CYCLING PARAMETERS

Phase 1	Step	Temp	Time
1	Initial denature	98°C	3°min
Stop and add 2.5 µL Phusion polymerase per tube			

Phase 2	Step	Temp	Time
2	Initial denature	98°C	20 s
3	Denature	98°C	10 s
4	TD anneal	70–55°C	30 s
5	Elongation	72°C	3′45″
Steps 3–5 repeated 15 times, temp decreasing 1°C/cycle			

Phase 3	Step	Temp	Time
6	Denature	98°C	10 s
7	Anneal	58°C	30 s
8	Elongation	72°C	3′45″
Steps 6–8 repeated 33 times			

Termination	Step	Temp	Time
9	Elongation	72°C	7 min
10	Halt reaction	11°C	10 min

The Phusion DNA polymerase is added after the initial 3°min denaturation step. In the touchdown phase is the annealing temperature is set 12–15°C above the theoretical annealing temperature of the primers and lowered by 1°C for 15 cycles. After this the annealing temperature is set to 58–60°C.

transformed cells plated out on ampicillin selective agarose. Colonies were picked the following day and grown in 500 µL of NZYM medium (10 g NZ amine, 5 g NaCl, 5 g yeast extract, and 2 g $MgSO_4.7H_2O$ in 1 L of ddH_2O, adjusted to pH 7.2) for 6 h, after which 1.5 mL of LB medium with 100 µg/mL ampicillin was added and grown overnight. Plasmid DNA was extracted using Nucleospin Miniprep kit and analyzed by restriction enzyme digests; positive samples were submitted to AGRF for sequencing.

Immunofluorescent confocal microscopy

Mandible osteoblasts (mOBs) grown in complete DMEM and human prostate cancer PC-3 cells (ATCC CRL-1435) grown in F12 HAMS with 10% fetal calf serum, were seeded onto glass coverslips and cultured for 72 h. After being fixed in 4% paraformaldehyde (PFA) and permeabilized with 0.1% Triton X the cells were blocked in 1% bovine serum albumin (BSA)/PBS for 1 h and then incubated with an antiC12orf29 rabbit polyclonal antibody (ab107423l; Abcam, Sapphire Biosciences) at 1:50 dilution for 1 h at room temperature. The cells were next labeled with an Alexa-Fluor 488 anti-rabbit secondary antibody (1:400; Life Technology) and counterstained with Rhodamine Phalloidin (1:1000; Life Technology) and DAPI (1:2000; Life Technology). The coverslips were mounted in Prolong Gold mounting medium. Confocal images were captured with a Leica TCS SP5 confocal laser scanning microscope (Leica Microsystems) fitted with 40× and 63× oil immersion objectives. Image analysis was performed with the Leica LAS

AF software package, and ImageJ (rsbweb.nih.gov/ij/) was used to pseudo color the images.

Immunohistochemistry

Tibiae and femora were harvested from 2-month-old Wistar rats, fixed in 4% PFA for 48 h, and then decalcified in 10% EDTA over a period of 4–5 weeks. The samples were embedded in paraffin and sectioned to 5 µm thickness with a microtome. Standard immunohistochemistry (IHC) protocols were applied and Proteinase K was used for antigen retrieval. The sections were incubated with an antiC12orf29 antibody (Abcam) at 1:75 dilution overnight at 4°C. After being washed, the slides were incubated with a DAKO biotinylated swine-anti-mouse, rabbit, goat secondary antibody (DAKO Australia Pty Ltd.) for 15 min, followed by 15 min incubation with horseradish perioxidase-conjugated avidin-biotin complex (DAKO Australia Pty Ltd.). Antibody complexes were visualized by the addition of a buffered diaminobenzidine. Safranin O staining was performed using a standard protocol.[13] Images were acquired with a Zeiss Axio Imager microscope (Carl Zeiss Pty. Ltd.) and analyzed using the ZEN image software.

Western blot analysis

Sheep mOBs and periodontal ligament (PDL) cells were grown in complete DMEM in a six-well plate for 3 days then harvested in a lysis buffer (10 mM Tris HCl pH 8, 150 mM NaCl, and 0.05% Tween) containing a protease inhibitor cocktail (Roche Diagnostics) after which the lysates were homogenized with a 23 g syringe and cell debris pelleted by centrifugation at 14,000 rpm for 20 min at 4°C. Clarified lysates were separated on a 10% SDS-PAGE gel and transferred to a nitrocellulose membrane by semi-dry transfer (Bio-Rad). The membranes were blocked in 2% BSA-TBS-Tween then incubated with the Abcam antiC12-orf29 antibody at 1:500 dilution then with a goat anti-rabbit secondary antibody (Thermo Scientific) and visualized by enhanced chemiluminesence and exposed on X-ray film (Fujifilm). A rabbit polyclonal anti-α-tubulin antibody (ab15246; Abcam) was used as a loading control and applied to the same membrane (1:2000 dilution) using the Licor Odyssey infrared system (Millenium Science).

Results

Result from isolation of randomly selected clones

A total of 30 randomly selected colonies were propagated in LB medium and the plasmid DNA extracted. Twenty-nine out of the 30 (96.6%) of the colonies contained cDNA inserts with an average size of 1141 nucleotides (nt), ranging from 268 to 3607 nt. Twenty of the clones (68%) had no existing sheep entries in GenBank and of these, the full coding sequence of 13 clones (65%) was identified. A complete mRNA sequence for *O. aries FTH1* was assembled from two separate clones: clone p7.46 contained 84% of the existing *O. aries FTH1* (NM_001009786.1), whereas the remaining 5′ end of the mRNA came from a 5′ *FTH1* sequence that had hybridized to a *TGF-β3* clone isolated using the MACH protocol. The assembled mRNA sequence (JX534528) contained an ORF that translated into 181 amino acids, which aligned with only 161 amino acids from

TABLE 3. LIST OF ALL CLONES ISOLATED FROM THE COMPLIMENTARY
DNA LIBRARY WITH GENBANK ACCESSION NUMBERS

Clone ID	Gene description	Gene name	Ovine mRNA in GB	Insert size	CDS coverage	GB accession number
p3.1	Stearoyl-CoA desaturase variant A	SCD1	no	1402	no	JX889613
p3.2	NIPA-like domain containing 3	NIPAL3	no	3607	no	JX889615
p3.3	Cbp/p300-interacting transactivator 2	CITED2	partial	1518	complete	JX534540
p3.4	Serpin peptidase inhibitor clade H member 1	SERPINH1	no	1813	complete	JX534537
p3.5	Proteasome (prosome, macropain) subunit, alpha type, 2	PSMA2	no	888	complete	JX534539
p3.6	Annexin A5	ANXA5	no	1729	partial	JX534527
p3.7	MAX dimerization protein 4	MXD4	no	1021	complete	JX534542
p4.1	Peptidylprolyl isomerase A	PPIA	partial	717	complete	JX534530
p4.2	Syntaxin 12	STX12	no	524	no	JX889612
p4.3	Peroxiredoxin 5	PRDX5	no	791	complete	JX889614
p4.4	Cystatin c	CST3	no	724	complete	JX534543
p4.5	Coiled-coil domain containing 80	CCDC80	no	1434	partial	JX534534
p4.6	Chromosome 12 open-reading frame 29	C12orf29	no	1176	complete	HQ438587
p4.7	Cytochrome c oxidase subunit 3	COIII	complete	754	partial	JX534526
p5.1	Ribosomal protein L17	RPL17	no	632	complete	JX534521
p5.2	Catenin (cadherin-associated protein), beta 1	CTNNB1	partial	817	partial	JX534532
p5.4	Maspardin isoform b	SPG21	no	933	partial	JX534523
p5.5	Binding protein S1, serine-rich domain	RNPS1	no	1855	complete	JX534545
p5.7	Vimentin	VIM	partial	1509	partial	JX534524
p6.1	CD9 molecule	CD9	complete	626	partial	JX534520
p6.2	Aminolevulinate, delta-, synthase 1	ALAS1	no	1700	partial	JX534541
p6.3	Kallikrein-related peptidase 7, pseudogene	KLK7 Ψ	no	670	N/A	JX534538
p6.4	Phosphatidylinositol glycan anchor biosynthesis, class F	PIGF	no	944	complete	JX534535
p6.5	Ribosomal protein S17	RPS17	no	484	complete	JX534531
p6.6	Homocysteine-inducible, endoplasmic reticulum stress-inducible, ubiquitin-like domain member 1	HERPUD1	no	762	partial	JX534522
p6.7	Heat shock 70kDa	HSP70	partial	268	no	JX534525
p7.1	Mitochondrial DNA D-loop	D-loop	complete	764	no	JX534533
p7.18	Ribosomal protein S8	RPS8	no	761	complete	JX534536
p7.39	Tumor protein, translationally controlled 1	TCTP	partial	834	complete	JX534529
p7.46	Ferritin, heavy polypeptide 1	FTH1	partial	929	complete	JX534528
ACTB_SLIP	Actin, beta	ACTB	complete	1905	complete	HM067830
GAPDH_MACH	Glyceraldehyde-3-phosphate dehydrogenase	GAPDH	partial	1285	complete	HM043737
TGF-β3_MACH	Transforming growth factor, beta 3	TGF-β3	partial	1862	partial	JX534544

the existing *O. aries* FTH1 protein (NP_001009786) but was 100% identical with the goat FTH1 protein sequence (ABL07498). NCBI has updated the GenBank version of *O. aries FTH1* to reflect this (NM_001009786.2). The longest complete mRNA sequence isolated from the library was *RNPS1* (JX534545), at 1855 nt, followed by *SERPINH1* (JX534537) at 1813 nt; whereas the shortest mRNAs with complete coding sequences were *RPL17* (JX534521) and *RPS17* (JX534531), at 632 nt and 484 nt respectively. The longest clone was *NIPAL3* (JX889615) at 3.6 kb and the shortest clone was *HSP70* at 268 nt (JX534525), both of which were 3′ untranslated regions (UTRs). GenBank accession numbers of the submitted sequences are listed in T3 ▶ Table 3.

Isolation of GAPDH using the MACH protocol

A full-length mRNA clone containing the gene *GAPDH* was isolated using the MACH isolation method adapted from Haerry and O'Connor.[11] The sequence was 1285 nt long and contained the full ORF, translating into a 333 amino acid, which was 99.4% identical (331/333) with *Bos taurus* GAPDH (NP_001029206). A BLASTP search revealed that *O. aries* GAPDH has a unique amino acid substitution variation at amino acid 23. In the vast majority of animals this residue is a serine (S), whereas in the sheep this residue is a threonine (T); the nomenclature for this variation is p.S23T.[14] This amino acid variation is caused by nonsynonymous single-nucleotide polymorphism (snSNP) on the first base of codon 23, which in most animals is *tct*; in the sheep this codon is *act* (c.67T > A). Serine and threonine are both small hydrophilic amino acids, which differ only in that threonine has a methyl group instead of a hydrogen group found in serine. This is therefore a conservative substitution since both amino acids can reside within the interior and on the surface of a protein.[15] The sequence of the isolated *GAPDH* clone was compared with partial *O. aries* mRNA sequences in GenBank (AF035421 and

AF272837) and an EST (Oa1229.1) from an *O. aries* EST library published by Hecht *et al.*[16] A threonine codon was found at this position in all three sequences, which suggests that this snSNP was not a sequencing artifact. The only other mammalian species that appears to share this serine to threonine substitution is the African bush elephant (*Loxodonta africana*) although this particular sequence was predicted by automated computational analysis from a genomic sequence (XM_003410789) and has therefore not been verified. The only other amino acid substitution in *O. aries* GAPDH protein compared to *B. taurus* GAPDH is an asparagine to histidine substitution at position 55 (p.N55H), which is caused by a first base adenine to cysteine substitution (c.163A > C). This amino acid substitution, unlike the p.S23T substitution, is not unique to sheep; other animals such as pig (NP_001193288), rabbit (XP_002708267) and horse (NP_001157328) have histidine residues at this position. The GAPDH mRNA sequence has the GenBank accession number HM043737.

Initial characterization of C12orf29

Clone p4.6 contained the full-length mRNA sequence of the gene *C12orf29* (accession number HQ438587), which translated into a 325 amino acid protein of unknown function. *C12orf29* is a highly conserved gene that is found in all vertebrate species that have been sequenced to date. A rabbit polyclonal antiC12orf29 antibody became available from Abcam within months of our isolation of the clone and this greatly aided the initial characterization of this gene. Confocal immunofluorescent (IF) microscopy showed that C12orf29 was highly expressed in mOBs cells (Fig. 3A), whereas the prostate cancer cell line PC3 has a low constitutive expression (Fig. 3B) in the same culture conditions. C12orf29 also appears to be expressed in the extracellular matrix (ECM) laid down by the mOBs (Fig. 3A, arrows), which suggests that the protein is exported out of the cells and is embedded within the ECM. Confocal z-stack analysis of individual cells showed no evidence of C12orf29 being expressed in the nucleus. Immunohistochemical staining of the tibial growth plate of the 2-month-old rats shows C12orf29 positive staining in the mineralization zone of the growth plate and in the adjacent trabecular bone (Fig. 4), coinciding with areas that stain brightly red for the glucosaminoglycan stain Safranin O (inset Fig. 4). Western blot analysis of cell lysates from mOBs and PDL cells from sheep show that there was a high constitutive expression of a 37 kDa protein (Fig. 5), which is in agreement with the predicted size for C12orf29. Subsequent blocking peptide experiments using the C12orf29 peptide used to generate the antibody showed a significant reduction of the intensity of the 37 kDa bands (data not shown).

Discussion

The first reports of cDNA libraries began to appear in 1975–1976[17–19] and the application of this procedure became commonplace during the 1980s, when reagents such as RNase H and DNA polymerase I (*Pol* I) simplified the process of second strand cDNA synthesis.[20]

The Stratagene pBluescript II XR cDNA library Construction Kit used in this study was primarily chosen on the basis that for a novice user, the relatively simple protocol of
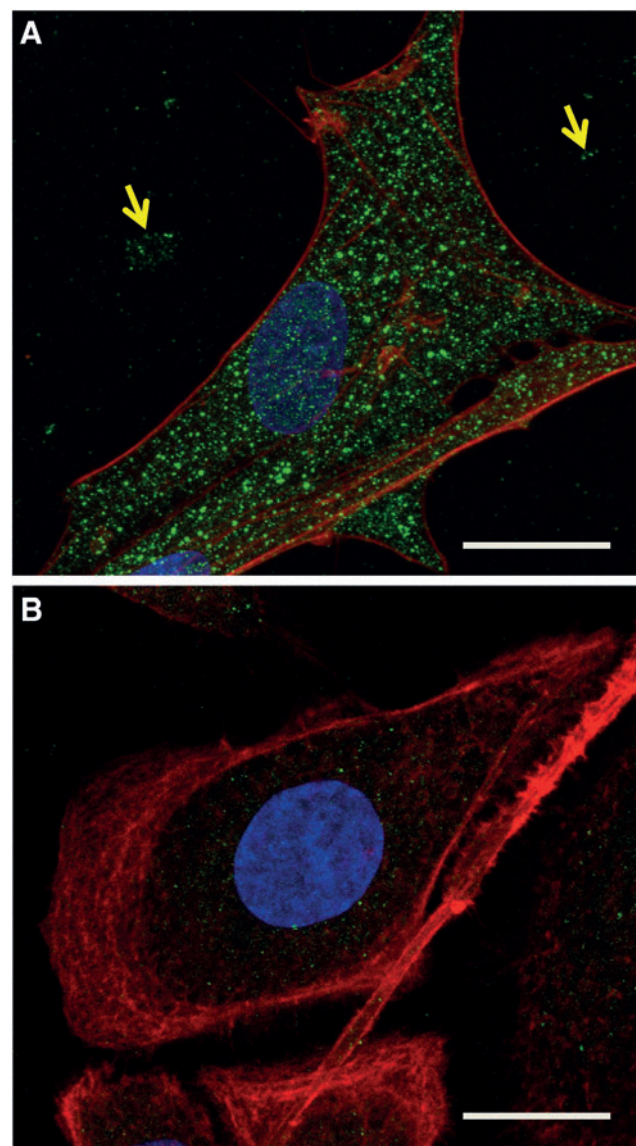

FIG. 3. Immunofluorescent confocal microscopy showing endogenous C12orf29 expression in mandible osteoblasts (mOB) and PC3 cells. Cells were labeled with anti-C12orf29 antibody and the expression visualized using an Alexa-Fluor 488-conjugated secondary antibody shown here in green. The cell nuclei are stained with DAPI (blue) and cytoskeleton is stained with Rhodamine Phallioidin (red). Scale bars = 20 μm. **(A)** A mOB cell showing strong C12orf29 staining throughout the cytosol and also in patches in the extracellular space (arrows). **(B)** A PC3 cell shows a faint but discernible C12orf29 staining within the cell. Color images available online at www.liebertpub.com/tec

this kit—a plasmid-based system—greatly increased the chance of a successful outcome. The chemistries used with the pBluescript kit are essentially the same as those first described by Gubler and Hoffman in 1983,[20] which was itself a modification of a system described by Okayama and Berg.[21] Prior to these publications, the limiting step in library construction was the reliance of the *Escherichia coli* DNA *Pol* I to use the initial reverse transcript both as a primer and template for second strand cDNA synthesis. The result was a hairpin double-stranded DNA with the 5′ end of
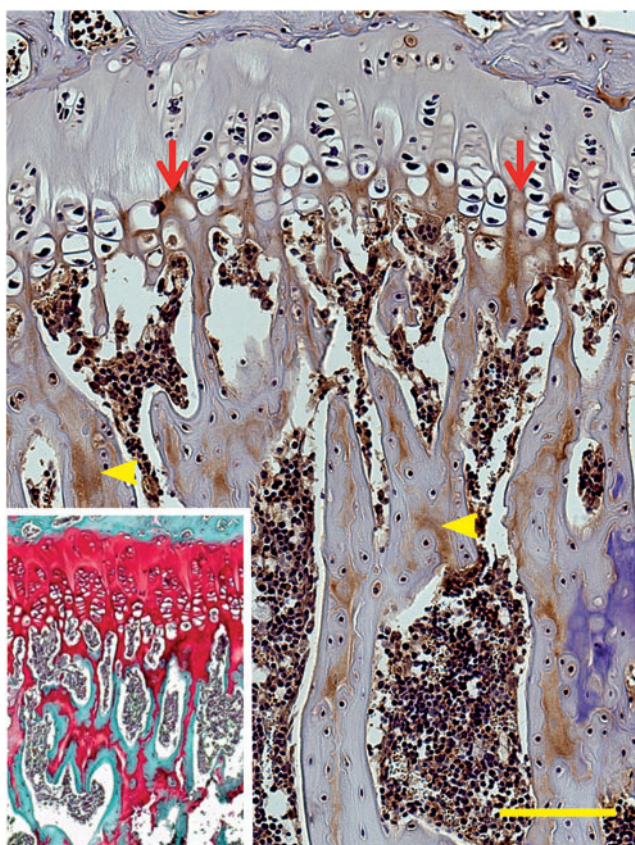
FIG. 4. Immunohistochemistry analysis of the tibial growth plate of a 2-month-old rat. Positive C12orf29 staining can be seen in the mineralized zone of the growth plate cartilage (red arrows) and also in the mineralized cartilage of the trabecular bone (arrowheads). These areas stain strongly red with Safranin O, evidence of high GAG/PG content (inset). Scale bar = 250 μm. Color images available online at www.liebertpub.com/tec

the mRNA in the single-stranded loop, which was cleaved with S1 nuclease digestive enzyme and resulted in loss of 5′ sequence information.[18] Once recombinant RNase H became available, it was then used in combination with *Pol* I
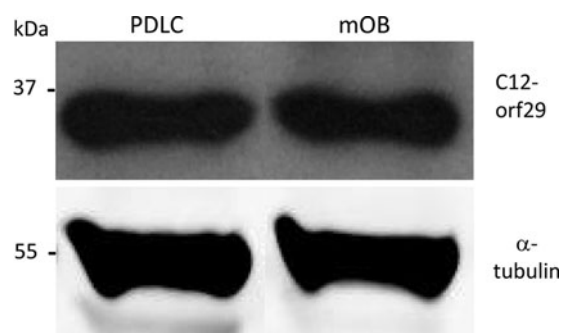

FIG. 5. Western blot analysis of whole cell lysates from mOBs and periodontal ligament cells (PDLCs). Immunoblotting show high C12orf29 protein expression in both these cell types. The band has a molecular weight of ~37 kDa, the predicted size of the 325 amino acid C12orf29 protein. An α-tubulin antibody was applied to the same membrane as a loading control.

for second strand synthesis; the RNase H creates nicks in the DNA:RNA hybrid strand, which then serves as a template for the *Pol* I enzyme.

Other cDNA library systems such as the Clontech SMART system offer advantages that make them a better choice for experienced users. Some of these benefits is the fact that the SMART system requires considerably less polyA$^+$ RNA template (1 μg vs. 5 μg for the pBluescript kit)—and if a Long-Distance PCR protocol is applied, as little as 50 ng of total RNA is sufficient; another benefit is that this system does not require methylation of the first strand as do other directional cloning methods, a process that is considered to be inefficient for cloning.[22] More importantly, the SMART system is designed to enrich for full-length cDNA by using an oligonucleotide that binds to the poly cysteine cap at the 5′ end of the first strand cDNA,[23] thereby effectively switching the reverse transcriptase template from the mRNA to the first strand cDNA. The parent template mRNA is removed by NaOH treatment, not RNase H as with conventional systems, and double-stranded cDNA is subsequently synthesized by thermostable DNA polymerase instead of *Pol* I DNA polymerase.[24]

Low versus high-throughput strategies

Two previous studies into the transcriptome of sheep bone had been performed using high-throughput EST library[16] and massively parallel DNA sequencing platform approaches; the latter is known as NGS.[25] Studies such as these generate large data sets that require high-performance computing infrastructure and specialized bioinformaticians to assemble and analyze. The study by Jager *et al.*,[25] for example, analyzed over 830,000 contigs with an average length of 134 bp using the preassembly program Velvet, which were then extended using the *de novo* transcriptome assembly program Oases. This two-step approach yielded a total of 117,594 extended contigs with an average length of 1374 bp. These contigs were BLAST searched against cow, mouse, and sheep mRNA transcripts and any contigs that could not be assigned to any existing transcripts were excluded from further analysis.

By contrast, since our goal was not to conduct a major transcriptome study, we opted for a low-throughput approach with an emphasis on isolating individual clones of interest, and we also characterize any novel genes identified from the random screening of colonies.

Library screening

Screening a large number of clones for GOI is a key step in many applications of recombinant DNA technology. Hybridization became the standard technique and is still widely used as the default screening method.[26] In this procedure the colonies are first lifted onto nitrocellulose filters placed on the surface of agarose plates, and a reference set of these colonies is made by replica plating and stored at 4°C. The colonies are lysed and their DNA denatured and fixed to the filter. RNA probes, typically ^{32}P-labeled, are then hybridized to the DNA and the fixed DNA with the GOI is located by autoradiography on X-ray film. The positive colonies are identified by aligning the X-ray film with the colonies on the original agarose plate.[27] Although a large number of colonies can thus be screened, the use of

nitrocellulose filters is a time-consuming task and tends to only work with highly expressed genes.[26]

PCR-based methods were developed in which cDNA clones were propagated as individual plaques on solid medium in 24-well plates and a phage suspension prepared from each well and transferred to a 96-well format—so-called arrayed PCR screening. This was followed by repeated PCR screening of incrementally smaller pools from which a positive signal had been observed. Although this approach did away with the need for filter hybridization and, more importantly, radioisotope labeling, the end result is a labor intensive and relatively insensitive method of screening.[28–30]

The MACH library screening method

Using the MACH protocol we were able to isolate several clones containing the full-length ORF of the housekeeping gene GAPDH from a cDNA library from cells derived from sheep bones. In addition, an 1862 nt sized clone containing the partial ORF of *O. aries* TGF-β3 was also isolated using this method (JX534544; data not shown). One of the main advantages of the MACH screening method is that it only requires the use of thermostable DNA polymerases and eliminates the need for costly and labile enzymes such as polynucleotide kinases and T4 DNA ligases. The linear PCR products form circularized plasmids by the action of base-pair complementarity of the overhangs that form when the two PCR products are combined, denatured, and annealed.[11] Thus, reverse strand 1 can only form a circularized product with forward strand 4, and likewise, the reverse strand 3 will only circularize with forward strand 2 (Fig. 2A, step 4). This mode of circularizing the plasmids is, in effect, a proof-reading feature that ensures the high fidelity of the insert, since it is highly unlikely that two strands of unequal length or homology would anneal and form compatible overhangs. We found it unnecessary to perform the phenol-chloroform DNA extraction following the annealing step that was recommended in the original article[11]; this eliminates the risk of losing the plasmid DNA in the washing steps.

C12orf29: a novel gene with a role in skeletal biology

Prior to our isolation of the *O. aries* C12orf29 clone, the full mRNA sequence of the gene had already been identified in a number of other vertebrates and mammals. As it happened, the mRNA had also been isolated in sheep by Jager *et al.*[25] but was wrongly annotated as CEP290, the 3′UTR of which, in all vertebrates, abuts the 3′UTR of C12orf29 on the opposite strand. The C12orf29 sequence from that study contained an atypical poly-adenylation signal (AATTAAA), similar to what we had identified in our clone; this feature was therefore unlikely to be an artifact. Their study does, however, provide an example that large data sets may obscure novel discoveries. C12orf29 immediately caught our attention as being the only gene in our relatively small sample of clones that had not been studied in any detail. The gene fitted the definition of a ''conserved hypothetical protein'' by having a wide phyletic distribution and unknown biochemical/physiological function.[31] The timely availability of C12orf29-specific antibodies allowed us to start characterizing its protein expression, and these data suggested that the protein may be a secreted ECM protein,

specific, but not necessarily limited, to skeletal cells and cartilage. IF microscopy showed a high protein expression in osteoblast cells (Fig. 3A), and in PDL cells and the pre-osteoblast cell line MC3T3-E1[32] (Supplementary Fig. S1A, B; ◀ SF1 Supplementary Data are available online at www.liebertpub.com/tec). A discernible fluorescent signal was also seen in patches between the highly expressing cells (arrows, Fig. 3A), indicating that C12orf29 protein was being embedded in the ECM. The strong protein expression seen in these cells was in stark contrast to that seen in the PC3 cells; however, these cells displayed what appears to be a low constitutive expression that could be seen in the cytosol (Fig. 3B). IHC analysis showed that C12orf29 was expressed in the growth plate of 2-month-old rats. The expression was found in the mineralized zone of the growth plate and also in the trabecular bone and appeared to overlap with areas that stained strongly for Safranin O in the calcified cartilage in this region; there was no positive C12orf29 staining observed in any demineralized bone sections (data not shown). Finally, western blot analysis of whole cell lysates showed that mOB and PDL cells both strongly expressed a protein that was detected at ∼37 kDa, the predicted molecular weight of the 325 amino acid protein encoded by mammalian *C12orf29* mRNA. Phylogenetic analyses show that C12orf29 has a pedigree that encompasses the width of the chordate superphylum.[33] This indicates a role that is highly conserved and, therefore, physiologically important. Our initial characterization strongly suggests that C12orf29 has role in the ECM of articular and growth cartilage and is, therefore, a novel structural protein in these skeletal tissues. We have yet to determine what post-translational modifications the protein undergoes, but we cannot rule out that it may be decorated with glucosaminoglycans and therefore potentially a proteoglycan.

Conclusions

We constructed a cDNA library from cells derived from sheep bone. The quality of the library was characterized and an initial screening of 30 random clones identified 20 genes that had no previous entries in the GenBank data base. The full-length cDNAs of the housekeeping genes *GAPDH* was isolated using the MACH protocol, described here in detail. From the randomly selected clones, we isolated a full-length cDNA of the uncharacterized gene *C12orf29*. The initial data concerning this gene suggest it has a role in skeletal biology and particularly in the ECM of cartilaginous tissues. Studies are currently underway to further characterize C12orf29.

Acknowledgments

Disclosure Statement

The authors have nothing to disclose.

References

1. Martini, L., Fini, M., Giavaresi, G., and Giardino, R. Sheep model in orthopedic research: a literature review. Comp Med **51,** 292, 2001.

2. Pearce, A.I., Richards, R.G., Milz, S., Schneider, E., and Pearce, S.G. Animal models for implant biomaterial research in bone: a review. Eur Cell Mater **13,** 1, 2007.

3. Muschler, G.F., Raut, V.P., Patterson, T.E., Wenke, J.C., and Hollinger, J.O. The design and use of animal models for translational research in bone tissue engineering and regenerative medicine. Tissue Eng Part B Rev **16,** 123, 2010.

4. Reichert, J.C., Saifzadeh, S., Wullschleger, M.E., Epari, D.R., Schutz, M.A., Duda, G.N., *et al.* The challenge of establishing preclinical models for segmental bone defect research. Biomaterials **30,** 2149, 2009.

5. Archibald, A.L., Cockett, N.E., Dalrymple, B.P., Faraut, T., Kijas, J.W., Maddox, J.F., *et al.* The sheep genome reference sequence: a work in progress. Anim Genet **41,** 449, 2010.

6. Das, M., Harvey, I., Chu, L.L., Sinha, M., and Pelletier, J. Full-length cDNAs: more than just reaching the ends. Physiol Genomics **6,** 57, 2001.

7. Nagaraj, S.H., Gasser, R.B., and Ranganathan, S. A hitchhiker's guide to expressed sequence tag (EST) analysis. Brief Bioinform **8,** 6, 2007.

8. Zhang, J., Chiodini, R., Badr, A., and Zhang, G. The impact of next-generation sequencing on genomics. J Genet Genomics **38,** 95, 2011.

9. Chaudhary, L.R., Hofmeister, A.M., and Hruska, K.A. Differential growth factor control of bone formation through osteoprogenitor differentiation. Bone **34,** 402, 2004.

10. Gregory, C.A., Gunn, W.G., Peister, A., and Prockop, D.J. An Alizarin red-based assay of mineralization by adherent cells in culture: comparison with cetylpyridinium chloride extraction. Anal Biochem **329,** 77, 2004.

11. Haerry, T.E., and O'Connor, M.B. Isolation of *Drosophila* activin and follistatin cDNAs using novel MACH amplification protocols. Gene **291,** 85, 2002.

12. Korbie, D.J., and Mattick, J.S. Touchdown PCR for increased specificity and sensitivity in PCR amplification. Nat Protoc **3,** 1452, 2008.

13. Rosenberg, L. Chemical basis for the histological use of safranin O in the study of articular cartilage. J Bone Joint Surg Am Vol **53,** 69, 1971.

14. den Dunnen, J.T., and Antonarakis, S.E. Nomenclature for the description of human sequence variations. Hum Genet **109,** 121, 2001.

15. Betts, M.J., and Russel, R.B. Amino Acid Properties and Consequences of Substitutions. Chichester, West Sussex, England; Hoboken, NJ: Wiley, 2003.

16. Hecht, J., Kuhl, H., Haas, S., Bauer, S., Poustka, A., Lienau, J., *et al.* Gene identification and analysis of transcripts differentially regulated in fracture healing by EST sequencing in the domestic sheep. BMC Genomics **7,** 172, 2006.

17. Rougeon, F., Kourilsky, P., and Mach, B. Insertion of a rabbit beta-globin gene sequence into an *E. coli* plasmid. Nucleic Acids Res **2,** 2365, 1975.

18. Efstratiadis, A., Kafatos, F.C., Maxam, A.M., and Maniatis, T. Enzymatic *in vitro* synthesis of globin genes. Cell **7,** 279, 1976.

19. Rougeon, F., and Mach, B. Stepwise biosynthesis *in vitro* of globin genes from globin mRNA by DNA polymerase of avian myeloblastosis virus. Proc Natl Acad Sci U S A **73,** 3418, 1976.

20. Gubler, U., and Hoffman, B.J. A simple and very efficient method for generating cDNA libraries. Gene **25,** 263, 1983.

21. Okayama, H., and Berg, P. High-efficiency cloning of full-length cDNA. Mol Cell Biol **2,** 161, 1982.

22. Zhu, Y.Y., Machleder, E.M., Chenchik, A., Li, R., and Siebert, P.D. Reverse transcriptase template switching: a SMART approach for full-length cDNA library construction. Biotechniques **30,** 892, 2001.

23. Xu, X., Vatsyayan, J., Gao, C., Bakkenist, C.J., and Hu, J. Sumoylation of eIF4E activates mRNA translation. EMBO Rep **11,** 299, 2010.

24. Clontech. SMART cDNA Library Constrcution Kit User Manual. PR7Y2399 ed2007. ◄AU3

25. Jager, M., Ott, C.E., Grunhagen, J., Hecht, J., Schell, H., Mundlos, S., *et al.* Composite transcriptome assembly of RNA-seq data in a sheep model for delayed bone healing. BMC Genomics **12,** 158, 2011.

26. Campbell, T.N., and Choy, F.Y. Approaches to library screening. J Mol Microbiol Biotechnol **4,** 551, 2002.

27. Grunstein, M., and Hogness, D.S. Colony hybridization: a method for the isolation of cloned DNAs that contain a specific gene. Proc Natl Acad Sci U S A **72,** 3961, 1975.

28. Alfandari, D., and Darribere, T. A simple PCR method for screening cDNA libraries. PCR Methods Appl **4,** 46, 1994.

29. Bloem, L.J., and Yu, L. A time-saving method for screening cDNA or genomic libraries. Nucleic Acids Res **18,** 2830, 1990.

30. Munroe, D.J., Loebbert, R., Bric, E., Whitton, T., Prawitt, D., Vu, D., *et al.* Systematic screening of an arrayed cDNA library by PCR. Proc Natl Acad Sci U S A **92,** 2209, 1995.

31. Galperin, M.Y., and Koonin, E.V. 'Conserved hypothetical' proteins: prioritization of targets for experimental study. Nucleic Acids Res **32,** 5452, 2004.

32. Wang, D., Christensen, K., Chawla, K., Xiao, G., Krebsbach, P.H., and Franceschi, R.T. Isolation and characterization of MC3T3-E1 preosteoblast subclones with distinct *in vitro* and *in vivo* differentiation/mineralization potential. J Bone Miner Res **14,** 893, 1999.

33. Friis, T.-E. The ancient gene C12orf29: an exploration of its role in the chordate body plan [Ph.D. Monograph]. Faculty of Science and Engineering, Queensland University of Technology, Brisbane, Australia, 2013.
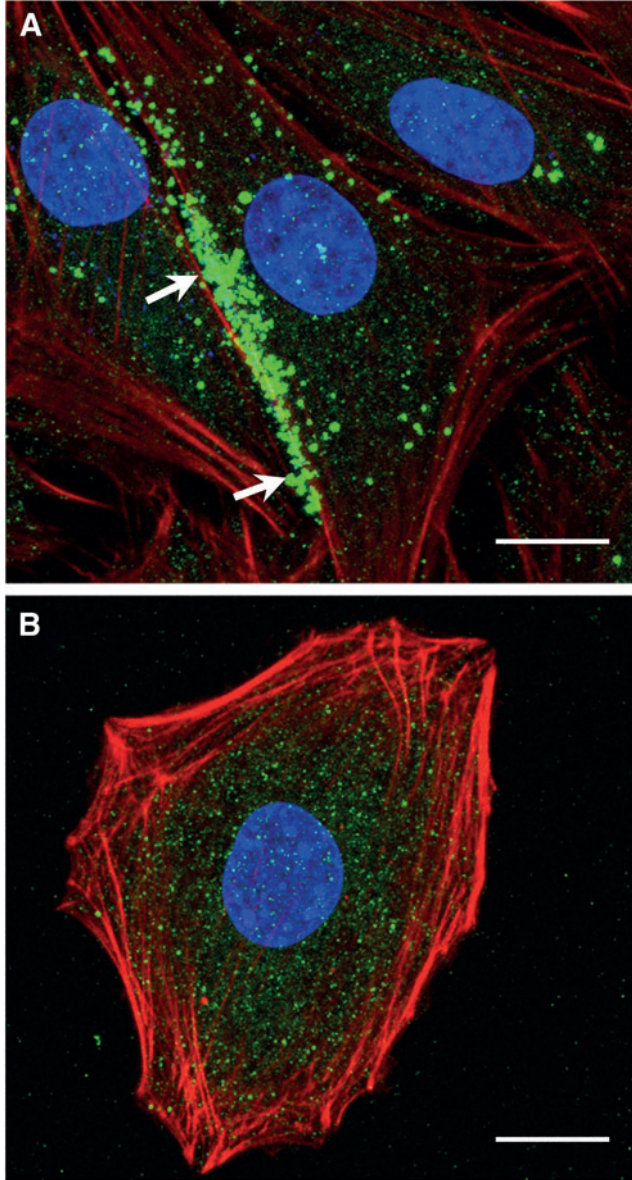
Address correspondence to:
Dietmar W. Hutmacher, PhD
Regenerative Medicine Group
Institute of Health and Biomedical Innovation
Queensland University of Technology
Kelvin Grove
Queensland 4059
Australia

E-mail: dietmar.hutmacher@qut.edu.au

Supplementary Data

SUPPLEMENTARY FIG. S1. Immunofluorescent confocal microscopy showing endogenous C12orf29 expression in sheep periodontal ligament and MC3T3-E1 cells. **(A)** Cells isolated from the periodontium of sheep show high expression of C12orf29 in the cytosol. The concentration is particularly strong along a ridge near the cell surface indicated by arrows. This feature is indicative of accumulation in the endoplasmic reticulum and that the protein is being exported out of the cells. **(B)** A MC3T3-E1 cell showing high C12orf29 expression in a punctate pattern within the cytosol. Scale bars = 20 μm.

AUTHOR QUERY FOR TEC-2013-0099-VER9-FRIIS_1P

AU1: Please note that gene symbols in any article should be formatted per the gene nomenclature. Thus, please make sure that gene symbols, if any in this article, are italicized.

AU2: Please review all authors' surnames for accurate indexing citations.

AU3: In Ref. 24, please mention the publisher's name and location.