Journal of proteome research

Article

pubs.acs.org/jpr

Analysis of Candidate Genes for Lineage-Specific Expression Changes in Humans and Primates

Cecilia Lindskog,*[,†] Martin Kuhlwilm,[‡] Armaity Davierwala,[‡] Ning Fu,[‡] Geeta Hegde,[§] Mathias Uhlén,[∥] Sanjay Navani,[§] Svante Pääbo,[‡] and Fredrik Pontén[†]

[†]Department of Immunology, Genetics and Pathology, Science for Life Laboratory, Uppsala University, Dag Hammarskjölds väg 20, SE-751 85 Uppsala, Sweden

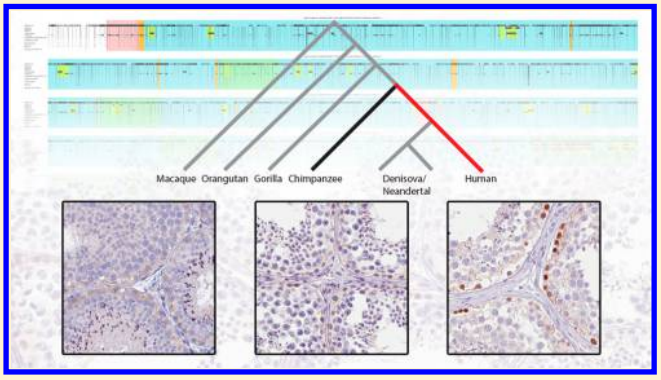
[‡]Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, 04103 Leipzig, Germany

[§]Lab Surgpath, 204 Bombay Market, 731/1 Tardeo Main Road, 400 034 Mumbai, India

[∥]Science for Life Laboratory, Royal Institute of Technology, Tomtebodavägen 23A, SE-171 21 Solna, Sweden

Supporting Information

ABSTRACT: *RUNX2*, a gene involved in skeletal development, has previously been shown to be potentially affected by positive selection during recent human evolution. Here we have used antibody-based proteomics to characterize potential differences in expression patterns of RUNX2 interacting partners during primate evolution. Tissue microarrays consisting of a large set of normal tissues from human and macaque were used for protein profiling of 50 RUNX2 partners with immunohistochemistry. Eleven proteins (AR, CREBBP, EP300, FGF2, HDAC3, JUN, PRKD3, RUNX1, SATB2, TCF3, and YAP1) showed differences in expression between humans and macaques. These proteins were further profiled in tissues from chimpanzee, gorilla, and orangutan, and the corresponding genes were analyzed with regard to genomic features. Moreover, protein expression data were compared with previously obtained RNA sequencing data from six different organs. One gene (*TCF3*) showed significant expression differences between human and macaque at both the protein and RNA level, with higher expression in a subset of germ cells in human testis compared with macaque. In conclusion, normal tissues from macaque and human showed differences in expression of some RUNX2 partners that could be mapped to various defined cell types. The applied strategy appears advantageous to characterize the consequences of altered genes selected during evolution.



KEYWORDS: *evolution, human, primate, lineage specificity, expression pattern, proteomics, immunohistochemistry, RNA sequencing, genomics*

INTRODUCTION

In a screen for positive selection in early modern humans, the Neandertal draft genome sequence was compared with that of five present-day humans from different parts of the world to look for large genomic regions where present-day humans share a common ancestor subsequent to their divergence from Neandertals. *RUNX2*, which was present within the top 20 of those regions,[1] is a transcription factor with target genes involved in skeletal development.[2] Mutations in *RUNX2* cause cleidocranial dysplasia, a condition characterized by delayed closure of cranial sutures, hypoplastic or aplastic clavicles, a bell-shaped rib cage, and dental abnormalities.[3] *RUNX2* also regulates the closure of the fontanel, which is essential for brain expansion[4] and controls the mesenchymal bone development that influences cranial morphology.[5] Recently Schlebusch et al.[6] have used an ancestral population branch statistics to identify candidates of positive selection in early modern humans and have shown that three of the top five positively selected regions contain genes involved in skeletal development, with *RUNX2* being one of the candidates.

Primate evolution has to date mainly been studied at the genomic and transcriptomic levels; however, studies are needed also at the proteomic level to fully understand the evolutionary changes between species. In particular, characterization of protein expression in various cell types across species is of major importance.

Tissue microarrays[7] (TMAs) and antibody-based proteomics[8] allow large-scale investigation of protein expression in histological samples. On the basis of this strategy, the Human Protein Atlas program (HPA) has been set up to map the human proteome. HPA generates affinity-purified antibodies and uses these for immunohistochemistry on TMAs from a large set of normal and cancer tissues as well as immuno-

Received: January 20, 2014
Published: June 9, 2014

dx.doi.org/10.1021/pr500045f | *J. Proteome Res.* 2014, 13, 3596−3606

fluorescence studies in human cell lines.[9−11] At present, data from more than 16 600 proteins are publically available at the Web site (http://www.proteinatlas.org), with extensive information on protein expression and validation of the antibodies.[12,13]

Because there are no coding differences between humans and archaic hominins,[14] but regulatory changes in the promoter,[15] we hypothesize that evolutionary changes in *RUNX2* expression and interactions affected aspects of the morphology of the upper body and cranium.[1] We used the HPA to further investigate *RUNX2* and its partners, defined as interacting partners and a manually curated list of activators, repressors, enhancers, and so on, and determined if there are differences in the expression patterns of these corresponding proteins between humans and apes. Furthermore, the availability of several great ape species allowed us to determine on which branch changes occurred.

From the original list of 66 *RUNX2* partners, 50 were selected for inclusion in the study based on antibody availability in the HPA and sequence identity to the macaque genome. The expression of the selected proteins was studied utilizing immunohistochemistry on formalin-fixed paraffin embedded tissues in a TMA format, including 32 different normal tissue types from humans and macaques. Proteins that differed in expression between the species were further studied on 27 different normal tissue types from chimpanzee, gorilla, and orangutan.

MATERIALS AND METHODS

Sample Collection

The human samples (three different individuals for each tissue) were received from Uppsala University Hospital, Sweden, as a part of the HPA.[10] Samples were collected from 32 different normal tissues, corresponding to 48 cell types, followed by formalin fixation and paraffin embedding. Samples from in total 187 different human individuals were used, originating from 85 females and 102 males. The average age was 58 ± 24 years, ranging from a 1 year old male (kidney and urinary bladder sample) to two 84 year old individuals (colon samples, one male and one female). Corresponding tissues from four adult rhesus macaque individuals (*Macaca mulatta*; Maq1: male, 5 years old; Maq2: female, 7 years old; Maq3: male, 6 years old; Maq4: female, 7 years old) were obtained from the German Primate Center, Göttingen, Germany. The macaques had been euthanized for other projects than the present study. In addition, tissue samples were prepared from in total 28 different normal tissues corresponding to 43 cell types from different species of apes. Two chimpanzees (Ch1: male, 3 days old; Ch2: female, 7 months old fetus) and two gorillas (Gor1: female, 2 h old; Gor2: female, 39 years old) were procured from Kolmården Animal Park, Norrköping, Sweden. One chimpanzee (Ch3: male, 40 days old) was obtained from Leipzig Zoo, Leipzig, Germany, and one orangutan (Orang: male, died during labor) was received from Furuvik Animal Park, Gävle, Sweden. Furthermore, a testis sample from one adult chimpanzee (12 years old) was obtained from BPRC, Rijswijk, The Netherlands, and used for immunohistochemical staining of *TCF3* in Figure 2A. All apes died of natural causes not related to this work.

Tissue Microarrays

The macaque and ape tissues were treated in the same manner as the human tissues prior to production of TMAs. The TMAs were generated essentially as previously described.[16] In brief, hematoxylin- and eosin-stained slides were histopathologically evaluated to select the appropriate area of sampling. Cylindrical cores with a diameter of 1 mm were removed from each donor block and placed into the recipient paraffin blocks. The macaque and ape tissues were, in general, included in duplicates from each donor block, while only one tissue core was used from each human individual. Nine TMAs consisting of in total 458 tissue cores were used in the present study. The cell types included in the TMAs are listed in Supplementary Table 1 in the Supporting Information.

Selection of Candidates for the Study

A list encompassing 66 candidate genes was compiled comprising partners of *RUNX2* (interacting partners and a manually curated list of activators, repressors, enhancers, etc.) Of these genes, 52 had antibodies available in the HPA at that time, including 28 in-house generated HPA antibodies and 24 commercially available antibodies from different companies. The protein epitope signature tag (PrEST) sequences used for generation of the HPA antibodies were further checked for identity to the macaque genome. Two of the 52 antibodies were excluded due to absence of sequence match toward the macaque genome or irrelevant gene with regard to available tissues (parathyroid hormone), resulting in a list of 50 genes included in the study (Supplementary Table 2 in the Supporting Information). The affinity-purified HPA antibodies used for the study were generated as previously described.[9,17,18] All antibodies that were generated as part of the HPA effort have undergone systematic steps for validation of functionality in various assays.[13] The quality assurance controls include analysis of protein arrays, Western blots, immunohistochemistry, and immunofluorescence in addition to comparing the outcome in each assay with published and bioinformatically predicted data when such is available (Supplementary Table 2 in the Supporting Information). In cases where more than one antibody was available for the same gene, the antibody with the best validation scores based on immunohistochemistry, immunofluorescence, Western blot, and protein array data was selected.

Immunohistochemistry

Immunohistochemistry and slide scanning were performed essentially as previously described.[16] In brief, 4 μm thick tissue sections from the TMA blocks were cut with a waterfall microtome (Leica Microsystems, Wetzlar, Germany) mounted on adhesive glass slides and baked for 45 min at 60 °C. Following deparaffinization in xylene and hydration in graded alcohols, blocking for endogenous peroxidase was performed in 0.3% hydrogen peroxide. For antigen retrieval, the slides were boiled for 4 min at 125 °C in a citrate buffer with pH 6 (Target Retrieval Solution, Dako, Glostrup, Denmark) using a pressure boiler (Decloaking chamber, Biocare Medical, Walnut Creek, CA). The dilution of primary antibodies was first optimized on test slides containing a small number of human tissues. Antibodies were diluted in UltraAb Diluent (Thermo Fisher Scientific, Fermont, CA), and automated immunohistochemistry was performed with an Autostainer XL ST5010 (Leica). In brief, the slides were incubated with the primary antibodies for 30 min at room temperature (RT), followed by detection with the secondary reagent antirabbit/mouse horseradish peroxidase-conjugated UltraVision (Thermo Fischer Scientific), applied for 30 min at RT. The slides were then developed for 10 min at RT, adding diaminobenzidine as a chromogen,

dx.doi.org/10.1021/pr500045f | J. Proteome Res. 2014, 13, 3596−3606

Table 1. Expression Differences between Macaque and Human on the protein level, Number of Cell Types Showing Expression and Number of Cell Types Differing in Expression between Species and between Individuals of the Same Species

gene name	cell types expressed macaque	cell types expressed human	cell types differing between macaque/human	cell types showing interindividual differences in macaque	cell types showing interindividual differences in human	cell types showing both interspecies and interindividual differences
AR	23	24	3	6	12	1
CDK1	41	42	5	6	22	0
CREBBP	48	42	6	3	17	1
EP300	48	48	4	10	10	0
FGF2	38	32	2	8	12	0
GLI3	39	17	19	11	8	5
HDAC3	46	46	6	19	5	0
HDAC5	46	46	12	33	24	10
JUN	40	34	7	16	20	4
PRKD3	34	33	5	19	9	1
RUNX1	26	19	3	11	6	1
SATB2	13	10	3	3	4	1
SMURF1	35	29	13	7	13	1
TCF3	2	14	7	0	1	0
XRCC5	45	44	2	7	9	0
YAP1	47	44	9	8	8	4

followed by counterstaining with Mayer's hematoxylin for 5 min (Sigma-Aldrich, St. Louis, MO) and mounting with Pertex (Histolab AB, Gothenburg, Sweden). All slides from the different species were treated simultaneously under equal conditions for every antibody.

Evaluation of Immunohistochemical Staining

The immunohistochemically stained and mounted slides were scanned with Aperio Scanscope XT (Aperio Technologies, Vista, CA), generating high-resolution digital images. The different cell types for every protein were manually scored as previously described.[19] In brief, intensity and fraction of immunostained cells was scored using four-graded scales. For intensity of staining: 0 = negative, 1 = low, 2 = medium, and 3 = high protein expression, and for fraction of positive cells: 0 = 0−1%, 1 = 2−10%, 2 = 11−25%, 3 = 26−50%, 4 = 51−75%, and 5 = 75−100%. A common score was set for duplicate cores on the TMAs representing samples from the same individual and organ. Slides were scored in a randomized manner, alternating between several antibodies of the same species before scoring the other species.

Computational Analysis of Genomic Features

Genomic features were analyzed using the catalog of differences between humans and the high coverage Denisova genome, containing alignments for chimpanzee, gorilla, orangutan, and macaque as well as information for the inferred ancestral genome for the great ape species.[20] Information about frequencies of alleles in human populations was retrieved from Ensembl variation resources[21] using the *biomaRt* package in R. Regulatory elements were used as defined in the Ensembl database (http://www.ensembl.org/info/docs/funcgen/regulatory_build.html). Data for histone modification, GERP scores, and UTRs were retrieved from Encode[22] and for transcription factor binding sites from TransFac/Biobase[23] via the UCSC browser using the R package *rtracklayer*.[24] The Transfac Matrix Database v7.0 provides matrices for binding motifs on transcription factors. These motifs are obtained from the literature. Plots were created using the R packages *gplots* and *lattice*. We estimated correlations between age and expression as Pearson correlations and *Z* scores as standard deviations from the mean across all correlations.

Comparison with RNA sequencing data of primate species was performed using Illumina sequencing files provided by the group of Henrik Kaessmann (Lausanne). This data set from Brawand et al.[25] contains RNA sequencing data of humans and macaques for six tissues (cerebral cortex, cerebellum, heart, kidney, liver, and testis). Only reads falling in constitutive exons for both species (unpublished data) were used for comparison. Reads were aligned to the human genome (version hg19, 1000 Genomes Project) or the *Macaca mulatta* genome (mmul1) using *Tophat* 2.0.6. Fragments per gene (defined as ENSEMBL identifiers) were counted using HTseq-count 0.5.3p3 (http://www-huber.embl.de/users/anders/HTSeq/doc/overview.html). For testing differential expression between human and macaque samples, the R/Bioconductor package DESeq 1.11.3[26] was used based on the negative binomial distribution. Lineage-specific RNA expression shifts based on a phylogenetic maximum-likelihood approach for the human lineage and between macaques and great apes (chimpanzees, bonobos, gorillas, and orangutans) were provided by Brawand et al.[25]

RESULTS

The 50 selected proteins were manually scored based on intensity of immunohistochemical staining and fraction of positive cells in 32 different normal tissues, corresponding to 48 cell types. In humans, 12 proteins were considered to be housekeeping (expressed in all cell types), 22 proteins showed ubiquitous expression (>75% of the cell types), 14 were moderately expressed proteins (25−75% of the cell types), and 2 proteins showed cell-type-specific expression (<25% of the cell types).

A difference of at least two steps in intensity or fraction levels between human and macaque in a certain cell type was required for proteins to be considered as potentially changed between species. Interindividual differences were taken into consideration; that is, the lowest individual value of the species with highest score was at least two steps higher than the highest individual value of the species with the lowest score. Differences in at least one cell type were observed for 42 of the 50 investigated proteins, in total 198 differences. The cell types with most the differences between species were neuronal cells in hippocampus (11 proteins) and germinal center cells in

dx.doi.org/10.1021/pr500045f | J. Proteome Res. 2014, 13, 3596−3606

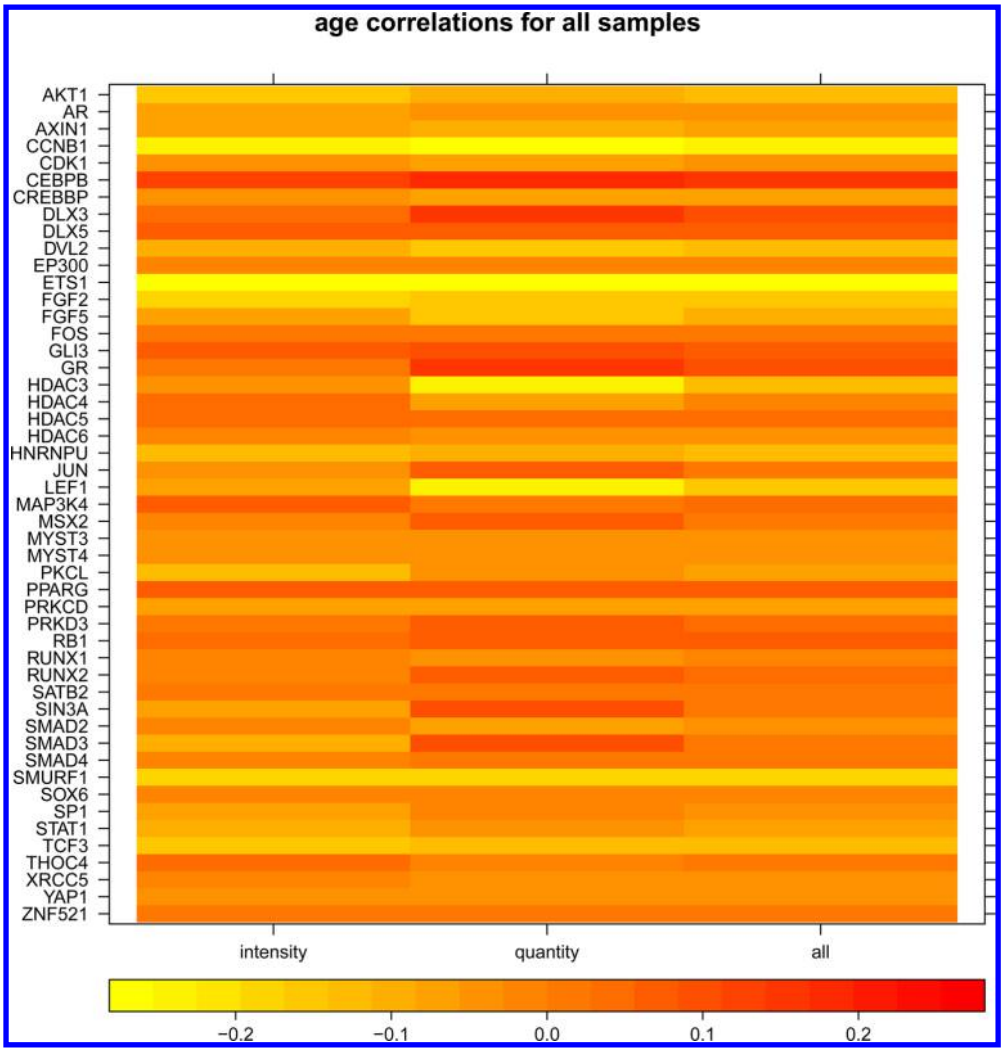


Figure 1. Heatmap of correlation between age and protein expression. Yellow indicates lower expression with age (negative correlation), while red indicates higher expression with age. Correlations are presented for intensity and quantity separately as well as merged across all data.

tonsil (11 proteins). The staining pattern in the cell types that differed was checked again manually, and 16 proteins displayed differences that seemed to reflect true changes in the protein expression between human and macaque (Table 1).

In addition to the changes observed between species, differences were also noted between individuals of the same species. For 47 of the 50 investigated proteins, interindividual differences in at least one cell type were found in macaques (in total 720 differences), while 49 proteins presented interindividual differences in at least one cell type in humans (604 differences). In macaques, most interindividual differences were observed in renal glomeruli (27 proteins) and glial cells in lateral ventricle (24 proteins), while humans mainly differed in alveolar macrophages (33 proteins) and neuronal cells in cerebral cortex (24 proteins). Table 1 shows the number of cell types with differences within the same species for the 16 proteins that changed between macaque and human. To investigate if differences within the same species were related to age, a heatmap figure was generated for all 50 proteins (Figure 1). Several proteins revealed a slightly higher intensity with age, but some proteins showed the opposite pattern. *Z* scores were calculated from standard deviations for these correlation sets (Supplementary Table 3 in the Supporting Information), indicating that CCNB1, ETS1 (down) and CEBPB (up) may change strongly in their expression pattern with age (*Z* score >2). These changes in expression with age were not detected as interspecies differences in this study.

Comparison of Protein Expression Data and RNA Sequencing Data between Macaque and Human

The observed expression differences seen between humans and macaques at the protein level were compared with RNA sequencing data for six different organs. After analyzing differential expression between human and macaque RNA samples, 1 out of the 16 genes was significantly different in both approaches: *TCF3* was upregulated in human testis ($p = 0.003$), which is in concordance with the protein expression data regarding tissue and direction of the expression difference.

Three genes (*CDK1*, *HDAC5*, and *RUNX1*) changed in expression in both data sets but not in the same tissues. Of the in total 50 genes, transcripts for 42 were available in the RNA sequencing data set. Transcription levels for 7 other of those 42 investigated genes (*FOS*, *PPARG*, *PRKCD*, *RUNX2*, *SOX6*, *SMAD3*, *STAT1*) passed the significance threshold ($p < 0.1$) in at least one tissue between humans and macaques (Table 2). However, corresponding protein expression differences in similar tissues were observed for none of them.

Furthermore, for *SATB2*, a higher expression at the transcript level was observed in macaque cerebral cortex compared with

dx.doi.org/10.1021/pr500045f | *J. Proteome Res.* 2014, 13, 3596−3606

Table 2. Expression Differences between Macaque and Human on the RNA Level, Fold Change between Human and Macaque, and Benjamini–Hochberg Corrected *p* Values

gene name	tissue	fold change	*p* value
CDK1	cerebellum	0.036	0.03
FOS	testis	6.073	0.02
HDAC5	cerebellum	2.763	0.08
PPARG	cerebellum	21.541	0.07
PPARG	testis	5.040	0.05
PRKCD	kidney	12.758	0.05
PRKCD	testis	10.193	0.07
RUNX1	cerebellum	0.182	0.09
RUNX2	cerebellum	0.032	0.0000009
SMAD3	testis	5.340	0.01
SOX6	cerebellum	0.295	0.08
SOX6	testis	0.197	0.06
STAT1	testis	3.798	0.03
TCF3	testis	8.788	0.003

human cerebral cortex but not significant after Benjamini–Hochberg correction for multiple testing (raw *p* value 0.03, adjusted *p* value 0.28). Although *SATB2* failed to reach formal significance, the 2.7 times higher expression in macaques could be biologically relevant. No obvious difference was observed between macaque and human cerebral cortex at the protein level; however, in hippocampus humans showed lower protein expression than macaques.

Protein Expression Analysis in Apes

The 16 proteins that were identified showing differences between human and macaque were further stained on TMAs from chimpanzee, gorilla, and orangutan, together with two out of the four macaque individuals and three new human individuals to check the reproducibility and to identify changes specific for a certain branch. After the second staining, 11 proteins showing in total 56 differences remained as changed between species, while five proteins could be ruled out. Four of those (CDK1, GLI3, HDAC5, and XRCC5) showed fewer differences between species compared with the first staining. CDK1, GLI3, and XRCC5 changed in expression mainly in the human tissues between the two stainings, which can be explained by either biological variation or technical variation. HDAC5 differed slightly in expression intensity in the macaque tissues between the two stainings, possibly explained by technical variation. One protein was ruled out due to differences in endothelial cells rather than in the organ-specific cell types (SMURF1).

The analysis of expression differences in great apes and macaques on the protein level enabled us to search for changes specific to the respective branch (Figure 2). Of the 11 proteins displaying expression differences, two (CREBBP and YAP1) showed cell-type-specific expression differences on both the human lineage and on the human–chimpanzee branch. This resulted in four genes that differed in expression between great apes and macaques, three genes that showed differences on the human–chimpanzee branch and six genes that differed on the human lineage after the split from chimpanzees (Table 3).

One gene was identified showing similar expression differences in both the TMA and RNA data sets (*TCF3*). The immunohistochemical staining of TCF3 in human testis revealed high expression in the basal cells (spermatogonia) of seminiferous ducts, while the corresponding cell type was weak or negative in chimpanzee and macaque (Figure 3A). In

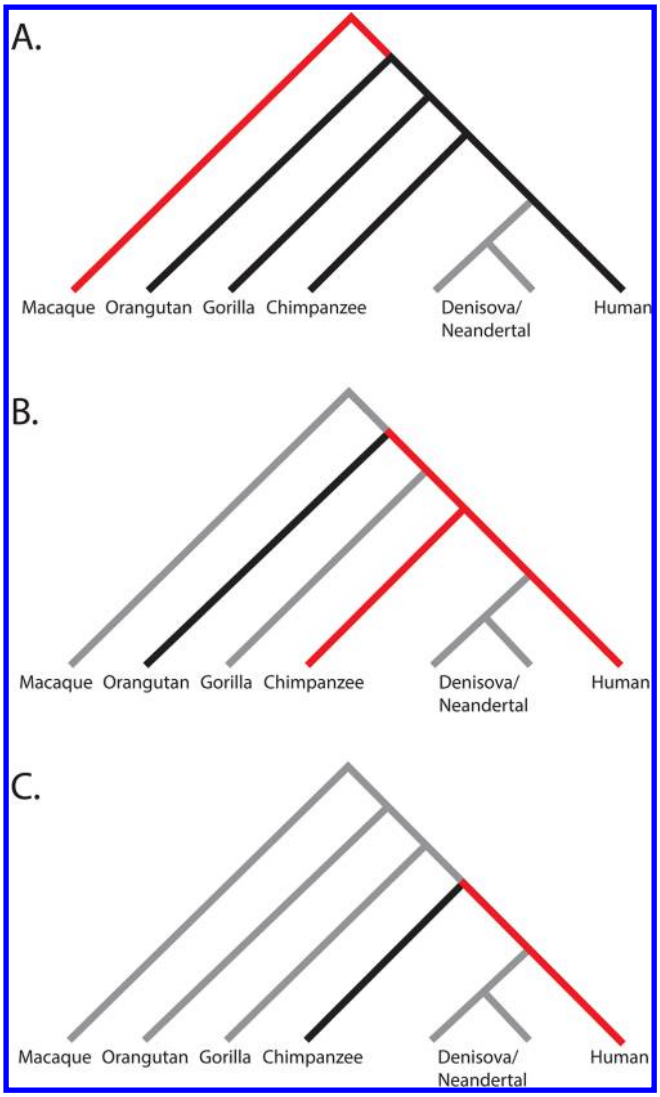


Figure 2. Phylogenetic tree of analyzed species. (A) Macaque versus great apes. (B) Orangutan versus human–chimpanzee lineage. (C) Human lineage versus chimpanzee. Red and black: Parts of the tree that were compared.

addition to the expression difference observed in testis, the TMA data showed high expression of TCF3 in a majority of the cells in germinal centra of human lymphoid tissues, while expression was found only in a small fraction of the germinal center cells in all other species (Figure 3B). *TCF3* was also among the genes detected to have changed in the RNA expression in testis on the human lineage,[25] taking the expression in the other great ape species into account.

SATB2 protein was partially expressed in human brain, with low or absent expression observed in hippocampus and higher nuclear expression observed in other parts of the brain. In chimpanzee, orangutan, and macaque, distinct nuclear expression was observed in neuronal cells in all parts of the brain (Figure 4A). Furthermore, the TMA data revealed that a lower fraction of both germinal center and nongerminal center lymphoid cells showed nuclear expression of SATB2 in humans compared with all other species (Figure 4B).

In addition to the changes previously described, the TMA data presented a number of distinct differences between species in the other nine genes (Table 2). Four of the proteins differed between macaque and great apes (AR, FGF2, HDAC3, and

dx.doi.org/10.1021/pr500045f | *J. Proteome Res.* 2014, 13, 3596–3606

Table 3. Genes That Differed in Expression in the TMAs in the Respective Lineages[a]

	gene name	Ensemble ID	tissues with different expression	observed differences
genes differing between great apes and macaques	*AR*	ENSG00000169083	adrenal gland	expression only in macaque
	FGF2	ENSG00000138685	bile ducts, pancreatic ducts	expression only in macaque
	HDAC3	ENSG00000171720	lymphoid tissue	lower expression in macaque
	PRKD3	ENSG00000115825	gastrointestinal tract	expression only in macaque
human−chimpanzee branch	*CREBBP*	ENSG00000005339	cerebellum, cerebral cortex	lower expression in human−chimpanzee
	RUNX1	ENSG00000159216	cerebral cortex, hippocampus, lateral ventricle	lower expression in human−chimpanzee
	YAP1	ENSG00000137693	cerebellum − molecular layer	lower expression in human−chimpanzee
human lineage-specific genes	*CREBBP*	ENSG00000005339	hippocampus	lower expression in human
	EP300	ENSG00000100393	hippocampus, lateral ventricle	lower expression in human
	JUN	ENSG00000177606	lateral ventricle	lower expression in human
	SATB2	ENSG00000119042	brain, lymphoid tissues	lower expression in human
	TCF3	ENSG00000071564	testis, lymphoid tissues	higher expression in human
	YAP1	ENSG00000137693	cerebellum − granular layer and Purkinje cells, cerebral cortex	lower expression in human

[a]Human−chimpanzee branch genes were tested using the inferred common ancestor with orangutan, while human lineage-specific genes were tested using the common ancestor with chimpanzee.

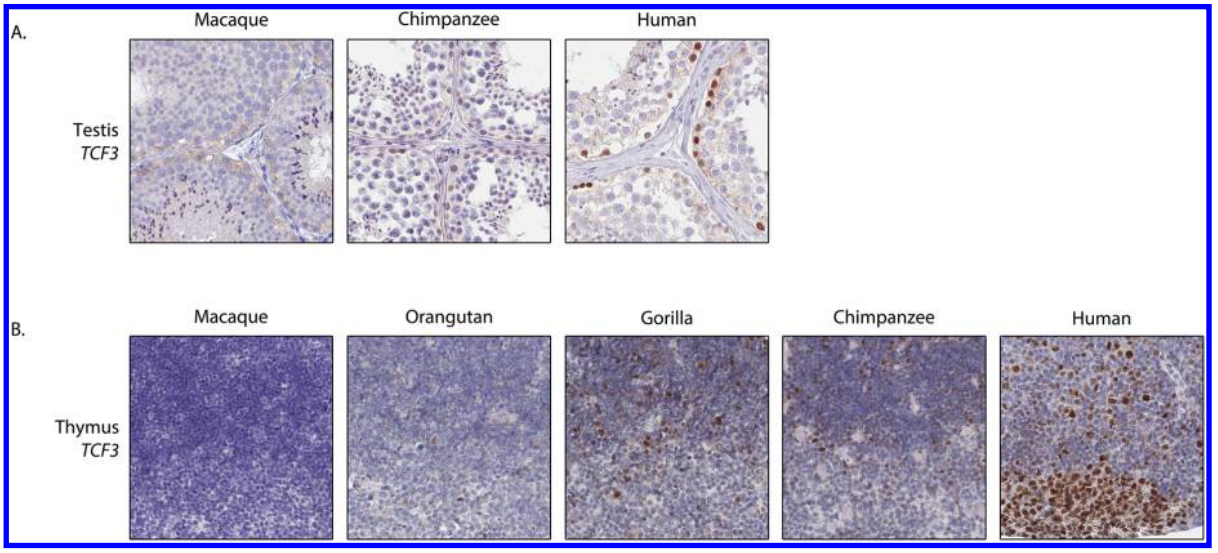


Figure 3. Immunohistochemical staining of TCF3 in testis and thymus, showing expression changes in the human lineage. Strong nuclear expression was observed in spermatogonia of human seminiferous ducts of testis (A) and in germinal centra of human thymus (B). The expression was lower in the other investigated species. The orangutan testis sample was not considered because it was not from an adult individual.

PRKD3; Supplementary Figure 1 in the Supporting Information). AR was highly expressed in a subset of cells in adrenal cortex (zona glomerulosa and zona fasciculata) in macaque, while no expression was observed in the corresponding cell types in any of the other species. FGF2 displayed moderate to high nuclear expression in pancreatic ducts and bile ducts of macaque; however, all ducts were negative in the other species. Furthermore, HDAC3 revealed high expression in both germinal center and nongerminal center lymphoid cells of all great apes, while low expression was observed only in a few cells in macaque. PRKD3 showed strong cytoplasmic expression in several epithelia of the gastrointestinal tract in macaque (appendix, colon, duodenum, and esophagus); however, the expression was lower or absent in the other species.

Five genes displayed differences in expression between species in different regions of the brain (Supplementary Figure 2 in the Supporting Information). One gene (*RUNX1*) showed high nuclear expression in brain of macaque and orangutan, while the expression was absent in the human-chimpanzee branch. On the RNA level, *RUNX1* showed a significantly lower expression in human cerebellum. Two genes, *CREBBP* and *YAP1*, revealed changes specific to both the human−chimpanzee branch and the human lineage, in different regions of the brain, with lowest expression observed in human brain. Furthermore, *EP300* and *JUN* showed moderate to high nuclear expression in the brain of all species except human.

Genomic Analysis

To analyze the 11 genes found to differ in protein expression on the genomic level, we searched for differences between the respective groups. This would allow us to identify putative candidate positions for the expression differences. Divergent positions between the great ape and the macaque branch were defined as positions that differed between macaque and the human reference, which had to be the same as the inferred

dx.doi.org/10.1021/pr500045f | J. Proteome Res. 2014, 13, 3596−3606

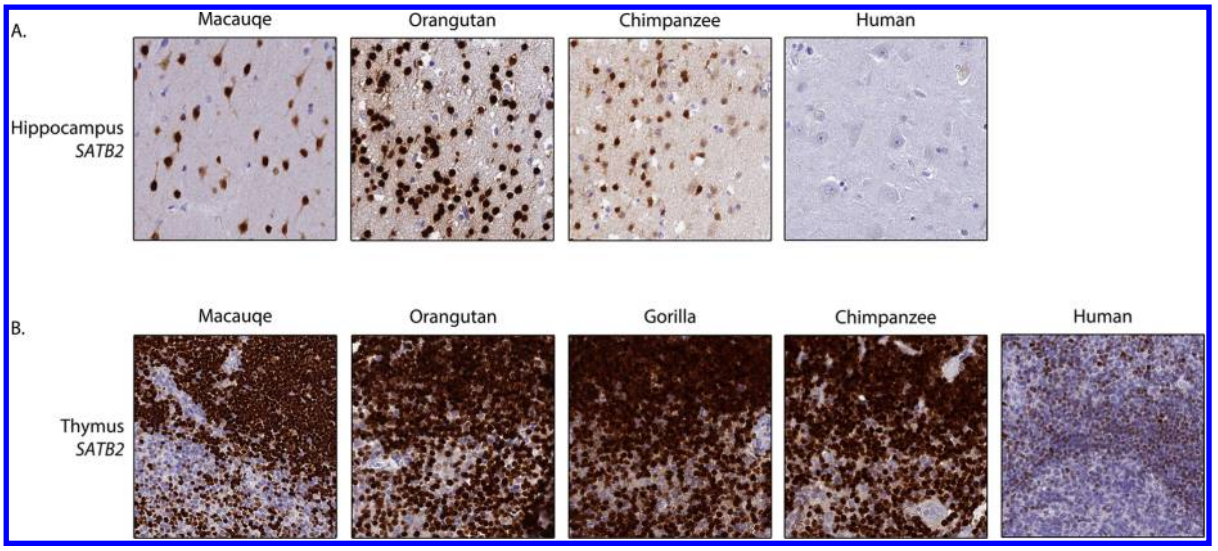


Figure 4. Immunohistochemical staining of SATB2 in hippocampus and thymus, showing expression changes in the human lineage. High nuclear expression was observed in a majority of both neuronal cells and glial cells in hippocampus of macaque, orangutan, and chimpanzee (A) as well as in both germinal center cells and nongerminal center cells in thymus of macaque, orangutan, gorilla, and chimpanzee (B). Human hippocampus was negative (A), while human thymus showed expression only in a subset of the cells (B).

Table 4. Number of Unique Positions, Substitutions and Indels That Differed between Lineages for the Tested Genes

	gene name	Ensemble ID	length	substitutions	indels	unique positions
genes differing between great apes and macaques	*AR*	ENSG00000169083	195996	6653	17503	24882
	FGF2	ENSG00000138685	81528	318	340	690
	HDAC3	ENSG00000171720	25994	1058	930	2102
	PRKD3	ENSG00000115825	84306	360	78	501
human−chimpanzee branch	*CREBBP*	ENSG00000005339	165672	2073	2027	4292
	RUNX1	ENSG00000159216	1206949	11188	10341	22586
	YAP1	ENSG00000137693	132962	1187	831	2179
human lineage-specific genes	*CREBBP*	ENSG00000005339	165672	925	427	1483
	EP300	ENSG00000100393	98291	455	443	979
	JUN	ENSG00000177606	13320	52	6	65
	SATB2	ENSG00000119042	211766	927	182	1239
	TCF3	ENSG00000071564	53033	369	941	1347
	YAP1	ENSG00000137693	132962	706	388	1218

common ancestor of humans and chimpanzees and either the gorilla or the orangutan ancestor (Figure 2A). Divergent positions on the human−chimpanzee branch were defined as positions that differed between the inferred common ancestor with orangutan and the human reference, and the human reference was the same as the common ancestor with chimpanzee (Figure 2B). Human lineage-specific differences were defined as positions where the human reference differed from the inferred common ancestor with chimpanzee (Figure 2C).

Analysis of Positions in Genes Showing Expression Differences

The positions (insertions/deletions and substitutions) that differed in the different lineages (Table 4) were identified and characterized. Information was collected about positions within putative promoters (i.e., regions within 5000 bp upstream of a transcription start site), Ensembl Regulatory Elements, exons or 5′ and 3′ untranslated regions (UTRs). GERP scores were retrieved, which reflect the conservation in mammals, transcription factor binding sites, histone modifications for promoter and enhancer activity (H3K4me1, H3K4me3, and H3K27Ac) and frequencies in human populations. Positions with a minor allele frequency in humans greater than 10% were excluded. For each gene that changed in expression in one of the branches, we looked at the putative promoters, regulatory elements, 5′ UTRs, 3′ UTRs, GERP scores (only scores >1 were considered), transcription factor binding sites, and histone modifications (only values >150 were considered). The positions were ranked based on the count of the previous criteria. The top positions in these lists could serve as candidates with the highest priority in further functional studies.

Plots were created with the divergent positions for the different species (human, inferred ancestor with chimpanzee, inferred ancestor with orangutan, inferred ancestor with gorilla, macaque, and Denisovans/Neandertals) and the other genomic features previously discussed (Supplementary Figure 3 in the Supporting Information). Again, this demonstrates the high number of divergent positions between the branches and the need to prioritize them.

Analysis of Transcription Factor Binding Sites

One of the features we characterized in the 11 genes that differed in protein expression was transcription factor binding sites (TFBS). Changes in TFBS can cause changes in gene

dx.doi.org/10.1021/pr500045f | J. Proteome Res. 2014, 13, 3596−3606

expression. If several genes change in expression on a certain branch, shared differences in TFBS would be of interest because these changes might reflect shared changes in their regulation. We analyzed the overlap in changed binding sites in the three subsets.

In the six genes that changed on the human lineage, 61 changed TFBS were observed, of which 5 TFBS were different in at least two genes. An analysis of the transcription factors binding to these five sites revealed that two were involved in bone/skeletal development/limb morphogenesis (*Brachury*, *PRRX2*), and three transcription factors (*Brachury*, *MEF2A*, and *POU6F1*) were involved in both heart/cardiac muscle development and neuronal development.

In the three genes that changed on the human−chimpanzee branch, 138 TFBS were changed, of which 27 were different in at least two genes. Analysis of the transcription factors binding to these 27 sites revealed that 5 were involved in bone/skeletal development/limb morphogenesis (*FOXL1*, *HNF1*, *JUN-D*, Myc, and *RUNX1*), 7 in apoptosis (*MEF2A*, *Myc*, *NKX2-5*, *PPARG*, *MECOM*, *TOPORS*, and *TP53*), 4 in heart/cardiac muscle development (*FOXL1*, *MEF2A*, *NKX2-5*, and *PPARG*), and 11 were involved in neuronal development (*ALX1*, *BPTF*, *FOXL1*, *IKZF1*, *MEF2A*, *NKX2-5*, *PPARG*, *RORA*, *MECOM*, *TGIF1*, and *TP53*).

In the four genes that changed in expression between great apes and macaques, 170 TFBS showed changes, of which 67 were different in at least two genes. Sixteen each were involved in apoptosis (*AHR*, *CREB1*, *E2F1*, *FOXO2*, *JUN*, *MAX*, *MECOM*, *MEF2A*, *MYC*, *NF1*, *NKX3-1*, *NR3C1*, *PAX4*, *STAT1*, *STAT5A*, and *TBP*) and neuronal development (ALX1, *Brachyury*, *BPTF*, *E2F1*, *EGR2*, *FOS*, *FOXL1*, *MECOM*, *MEF2A*, *NF1*, *NR3C1*, *PAX5*, *POU3F2*, *POU6F1*, *RORA*, and *STAT3*), and seven each were involved in heart/cardiac muscle development (*AHR*, *Brachyury*, *FOXL1*, *MEF2A*, *NF1*, *NKX3-1*, and *POU6F1*) and bone/skeletal development/limb morphogenesis (*Brachyury*, *HNF1A*, *HOXA9*, *MECOM*, *MYC*, *NF1*, and *PRRX2*.

DISCUSSION

To date, evolutionary studies have mainly been performed at the genomic or transcriptomic levels and not at the protein level. To our knowledge, this is the first time immunohistochemistry and TMAs have been used to analyze the evolutionary differences in protein expression between humans and other primates. The advantage of using antibody-based proteomics is that it is possible to look at the cell types individually without damaging the tissue composition, in contrast with RNA sequencing, where the entire tissue is homogenized and used for extraction. While RNA sequencing experiments allow a quantitative measurement of the abundance of transcripts, antibody-based proteomics provides information about the spatial distribution of proteins within a tissue and could give an estimate about the real abundance of the proteins. Furthermore, it has been found that only ∼40% of the variability observed in protein levels can be explained by mRNA levels,[27] and thus protein expression analysis provides a new level of information about the abundance of gene products in tissues. These abundance levels cannot be inferred from the genomic sequence alone, but retrieving more information about expression changes between tissues and species using different approaches could help in understanding how genomic information is related to expression. This comparative approach might even bear medical potential because it gives insight into phenotypes that are related to human diseases and could reveal functional consequences of genomic changes.[28]

Comparison of the protein expression data with the RNA sequencing data revealed similarities between the data sets, mainly for one gene, *TCF3*, showing significant differences between humans and macaques in the RNA sequencing data. In addition, this was the only gene that changed in expression on the human lineage in both the RNA sequencing data set and the TMA data set. Because we could confirm this observation on the protein level, this gene is a candidate for a selectively driven expression change in humans.

TCF3 (transcription factor 3) is a transcriptional regulator, involved in the initiation of neuronal differentiation. It is also involved in determination of tissue-specific fate during embryogenesis and is required for B lymphocyte development.[29] At the protein level, high expression was observed in human spermatogonia and lymphoid germinal center cells, while the corresponding cell types showed lower expression in all other species. The difference revealed in lymphoid tissues should be further characterized by investigating the distribution of the B and T lymphocytes across species, that is, using double staining with antibodies toward *TCF3* and different subtypes of lymphoid cells.

Although not significantly different on the RNA level, *SATB2* (SATB homeobox 2) could be an interesting candidate as well. This is a transcription factor that acts as a molecular node in a transcriptional network regulating skeletal development and osteoblast differentiation.[30] Leoyklang et al. have shown that a mutation in this gene leads to a phenotype that includes craniofacial dysmorphism and mental retardation.[31] Furthermore, the family of proteins that *SATB2* belongs to seems to play a role in the development of B lymphocytes and acute myeloid leukemia.[32] In addition to the role of *SATB2* in development, *SATB2* is selectively expressed in the lower gastrointestinal tract as well as in corresponding cancers of colorectal origin. This restricted expression pattern renders a potential role for SATB2 as a cancer biomarker.[33]

At the protein level, *SATB2* showed a similar level of expression in the lower gastrointestinal tract but lower expression levels in neuronal cells of both human hippocampus and lymphoid tissues compared with the other species. However, the protein was distinctly expressed in other parts of the human brain. Further studies are needed to deeper characterize the distribution of *SATB2* expression in different regions of the brain as well as in different subtypes of lymphoid cells. In particular, the observation that the craniofacial development is regulated by SATB2 in a dosage-sensitive way[34] and that we detect differences in expression in humans makes it a candidate for a human-specific functional change. The understanding of disease phenotypes caused by this gene could be improved by the evolutionary approach to measure phenotypes not only in the mouse model but also by variation in primates.

In four of the investigated genes (*AR*, *FGF2*, *HDAC3*, and *PRKD3*), distinct differences were observed at the protein level between macaques and the great apes. Macaque was used as an out-group in the genomic analyses, but no out-group for nonprimates has been used in our analysis. Thus, it is not possible to distinguish ancestral or derived states on either the macaque or the great ape lineage. We cannot determine if the observed differences occurred on the macaque branch or are differing between great apes, macaques, and their common ancestor.

For several of the investigated genes, differences in protein expression between species were displayed in various regions of the brain. Furthermore, differences including high expression in macaque and low expression in human were more common than the opposite. It is unclear if all observed differences reflect true changes in protein expression or if some of them might be explained by artifacts, possibly due to differences in handling of the tissues prior to generation of TMAs, that is, the length of time the tissues have been fixed in formalin.

Due to limited sample size, it is not possible to truly estimate the protein expression range in a population across species, and further studies on samples from more individuls are necessary to yield more robust data. Because only a small number of individuals were included for each species, the observed differences between species at the protein level could possibly be due to interindividual differences. Almost all investigated proteins showed interindividual differences in at least one cell type; however, a few of these differences were observed in the same cell types that differed between species. It should also be noted that changes observed between species despite interindividual differences in a certain cell type were strong changes because they were recorded between the highest individual score in one species and the lowest individual score in the other species. The second stainings on TMAs with the different apes also included three new human individuals, together with repeated staining on two of the macaque individuals, to check the staining reproducibility and to investigate the interindividual differences further. The proteins displaying distinct differences between the two species in both stainings are more reliable candidates for true changes in gene expression. Another characteristic that should be taken into account is the age of the investigated individuals. The human tissue samples are indeed from older individuals than the tissues from the other species, considering the life expectancy, and interindividual differences between humans proved to occur more often at higher age. This could influence the observed expression in certain cell types for proteins altered between preadult and adult individuals. The expression of two proteins (CCNB1 and ETS1) was inversely correlated with age. This could possibly be due to their function in osteoblast differentiation and periodontal ligament proliferation, which is presumably more important at younger age.[35] However, in an in-depth analysis of the protein expression in humans in relation to age, none of the proteins that changed between species showed a strong difference in expression pattern with age, and hence we assume that the effect of age differences is not a strong factor in our observations.

Beyond the findings for living species provided in this study, the high coverage Denisova genome allows us to look at more recent changes on the human lineage. If a regulatory position shows the ancestral state in this extinct hominin and a derived state in modern humans, one can speculate that the putative regulatory effect has arisen after the split of humans and Denisovans. In *SATB2*, which is significantly different in expression on the human lineage on the protein level, the genomic difference with the highest score shows that pattern. Thus, it is possible that the difference in expression of this gene might have occurred very recently in human evolution. In *JUN*, 3 out of 15 positions with the highest score were ancestral in Denisovans and derived in modern humans. These positions might also point toward a regulatory change in the modern human lineage. This hypothesis could be tested in further studies by using in vitro promoter assays that assess the influence of these differences on gene expression or by genome editing in human cells.

■ ASSOCIATED CONTENT

Supporting Information

Supplementary Figure 1. Immunohistochemically stained images with protein expression differences between great apes and macaques. Supplementary Figure 2. Immunohistochemically stained images with protein expression differences observed in different regions of the brain. Supplementary Figure 3. Gene map of *SATB2*. Differences between the human reference and the inferred ancestor of humans and chimpanzees, and their overlap with genomic features. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: Cecilia.Lindskog@igp.uu.se. Tel: +46 18 471 44 83. Fax: +46 18 55 33 54

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The members of the Human Protein Atlas (HPA) staff in Uppsala and Stockholm are acknowledged for their work. The project was financially supported by The Wallenberg Research Foundation (KAW) and the Max Planck Society. Tissue samples from chimpanzee, gorilla and orangutan were kindly provided and prepared by Tomas Bergström, SLU, Uppsala, and Katarina Eriksson, Pathology Department, Uppsala University Hospital. Ronald Bontrop, BPRC, Rijswijk is acknowledged for providing the adult testis tissue sample. Martin Kircher, University of Washington, is acknowledged for providing GTF files of constitutive exons.

■ REFERENCES

(1) Green, R. E.; Krause, J.; Briggs, A. W.; Maricic, T.; Stenzel, U.; Kircher, M.; Patterson, N.; Li, H.; Zhai, W.; Fritz, M. H.; Hansen, N. F.; Durand, E. Y.; Malaspinas, A. S.; Jensen, J. D.; Marques-Bonet, T.; Alkan, C.; Prufer, K.; Meyer, M.; Burbano, H. A.; Good, J. M.; Schultz, R.; Aximu-Petri, A.; Butthof, A.; Hober, B.; Hoffner, B.; Siegemund, M.; Weihmann, A.; Nusbaum, C.; Lander, E. S.; Russ, C.; Novod, N.; Affourtit, J.; Egholm, M.; Verna, C.; Rudan, P.; Brajkovic, D.; Kucan, Z.; Gusic, I.; Doronichev, V. B.; Golovanova, L. V.; Lalueza-Fox, C.; de la Rasilla, M.; Fortea, J.; Rosas, A.; Schmitz, R. W.; Johnson, P. L.; Eichler, E. E.; Falush, D.; Birney, E.; Mullikin, J. C.; Slatkin, M.; Nielsen, R.; Kelso, J.; Lachmann, M.; Reich, D.; Paabo, S. A draft sequence of the Neandertal genome. *Science (Washington, DC, U. S.)* **2010**, *328* (5979), 710−722.

(2) Kuhlwilm, M.; Davierwala, A.; Paabo, S. Identification of putative target genes of the transcription factor RUNX2. *PLoS One* **2013**, *8* (12), e83218.

(3) Mundlos, S.; Otto, F.; Mundlos, C.; Mulliken, J. B.; Aylsworth, A. S.; Albright, S.; Lindhout, D.; Cole, W. G.; Henn, W.; Knoll, J. H.; Owen, M. J.; Mertelsmann, R.; Zabel, B. U.; Olsen, B. R. Mutations involving the transcription factor CBFA1 cause cleidocranial dysplasia. *Cell* **1997**, *89* (5), 773−779.

(4) Falk, D.; Zollikofer, C. P.; Morimoto, N.; Ponce de Leon, M. S. Metopic suture of Taung (Australopithecus africanus) and its implications for hominin brain evolution. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109* (22), 8467−8470.

(5) Zhang, S.; Xiao, Z.; Luo, J.; He, N.; Mahlios, J.; Quarles, L. D. Dose-dependent effects of Runx2 on bone development. *J. Bone Miner. Res.* **2009**, *24* (11), 1889−1904.
(6) Schlebusch, C. M.; Skoglund, P.; Sjodin, P.; Gattepaille, L. M.; Hernandez, D.; Jay, F.; Li, S.; De Jongh, M.; Singleton, A.; Blum, M. G.; Soodyall, H.; Jakobsson, M. Genomic Variation in Seven Khoe-San Groups Reveals Adaptation and Complex African History. *Science* **2012**, 374−379.
(7) Kononen, J.; Bubendorf, L.; Kallioniemi, A.; Barlund, M.; Schraml, P.; Leighton, S.; Torhorst, J.; Mihatsch, M. J.; Sauter, G.; Kallioniemi, O. P. Tissue microarrays for high-throughput molecular profiling of tumor specimens. *Nat. Med.* **1998**, *4* (7), 844−847.
(8) Uhlen, M.; Ponten, F. Antibody-based proteomics for human tissue profiling. *Mol. Cell. Proteomics* **2005**, *4* (4), 384−393.
(9) Nilsson, P.; Paavilainen, L.; Larsson, K.; Odling, J.; Sundberg, M.; Andersson, A. C.; Kampf, C.; Persson, A.; Al-Khalili Szigyarto, C.; Ottosson, J.; Bjorling, E.; Hober, S.; Wernerus, H.; Wester, K.; Ponten, F.; Uhlen, M. Towards a human proteome atlas: high-throughput generation of mono-specific antibodies for tissue profiling. *Proteomics* **2005**, *5* (17), 4327−4337.
(10) Uhlen, M.; Bjorling, E.; Agaton, C.; Szigyarto, C. A.; Amini, B.; Andersen, E.; Andersson, A. C.; Angelidou, P.; Asplund, A.; Asplund, C.; Berglund, L.; Bergstrom, K.; Brumer, H.; Cerjan, D.; Ekstrom, M.; Elobeid, A.; Eriksson, C.; Fagerberg, L.; Falk, R.; Fall, J.; Forsberg, M.; Bjorklund, M. G.; Gumbel, K.; Halimi, A.; Hallin, I.; Hamsten, C.; Hansson, M.; Hedhammar, M.; Hercules, G.; Kampf, C.; Larsson, K.; Lindskog, M.; Lodewyckx, W.; Lund, J.; Lundeberg, J.; Magnusson, K.; Malm, E.; Nilsson, P.; Odling, J.; Oksvold, P.; Olsson, I.; Oster, E.; Ottosson, J.; Paavilainen, L.; Persson, A.; Rimini, R.; Rockberg, J.; Runeson, M.; Sivertsson, A.; Skollermo, A.; Steen, J.; Stenvall, M.; Sterky, F.; Stromberg, S.; Sundberg, M.; Tegel, H.; Tourle, S.; Wahlund, E.; Walden, A.; Wan, J.; Wernerus, H.; Westberg, J.; Wester, K.; Wrethagen, U.; Xu, L. L.; Hober, S.; Ponten, F. A human protein atlas for normal and cancer tissues based on antibody proteomics. *Mol. Cell. Proteomics* **2005**, *4* (12), 1920−1932.
(11) Lundberg, E.; Uhlen, M. Creation of an antibody-based subcellular protein atlas. *Proteomics* **2010**, *10* (22), 3984−3996.
(12) Uhlen, M.; Oksvold, P.; Fagerberg, L.; Lundberg, E.; Jonasson, K.; Forsberg, M.; Zwahlen, M.; Kampf, C.; Wester, K.; Hober, S.; Wernerus, H.; Bjorling, L.; Ponten, F. Towards a knowledge-based Human Protein Atlas. *Nat. Biotechnol.* **2010**, *28* (12), 1248−1250.
(13) Asplund, A.; Edqvist, P. H.; Schwenk, J. M.; Ponten, F. Antibodies for profiling the human proteome-The Human Protein Atlas as a resource for cancer research. *Proteomics* **2012**, *12* (13), 2067−2077.
(14) Castellano, S.; Parra, G.; Sánchez-Quinto, F. A.; Racimo, F.; Kuhlwilm, M.; Kircher, M.; Sawyer, S.; Fu, Q.; Heinze, A.; Nickel, B.; Dabney, J.; Siebauer, M.; White, L.; Burbano, H. A.; Renaud, G.; Stenzel, U.; Lalueza-Fox, C.; de la Rasilla, M.; Rosas, A.; Rudan, P.; Brajković, D.; Kucan, Ž.; Gušic, I.; Shunkov, M. V.; Derevianko, A. P.; Viola, B.; Meyer, M.; Kelso, J.; Andrés, A. M.; Pääbo, S. Patterns of coding variation in the complete exomes of three Neandertals. *Proc. Natl. Acad. Sci.* **2014**, DOI: 10.1073/pnas.1405138111.
(15) Doecke, J. D.; Day, C. J.; Stephens, A. S.; Carter, S. L.; van Daal, A.; Kotowicz, M. A.; Nicholson, G. C.; Morrison, N. A. Association of functionally different RUNX2 P2 promoter alleles with BMD. *J. Bone Miner. Res.* **2006**, *21* (2), 265−273.
(16) Kampf, C.; Olsson, I.; Ryberg, U.; Sjostedt, E.; Ponten, F. Production of tissue microarrays, immunohistochemistry staining and digitalization within the human protein atlas. *J. Vis. Exp.* **2012**, No. 63, 3620.
(17) Agaton, C.; Galli, J.; Hoiden Guthenberg, I.; Janzon, L.; Hansson, M.; Asplund, A.; Brundell, E.; Lindberg, S.; Ruthberg, I.; Wester, K.; Wurtz, D.; Hoog, C.; Lundeberg, J.; Stahl, S.; Ponten, F.; Uhlen, M. Affinity proteomics for systematic protein profiling of chromosome 21 gene products in human tissues. *Mol. Cell. Proteomics* **2003**, *2* (6), 405−414.
(18) Lindskog, M.; Rockberg, J.; Uhlen, M.; Sterky, F. Selection of protein epitopes for antibody production. *Biotechniques* **2005**, *38* (5), 723−727.
(19) Bjorling, E.; Lindskog, C.; Oksvold, P.; Linne, J.; Kampf, C.; Hober, S.; Uhlen, M.; Ponten, F. A web-based tool for in silico biomarker discovery based on tissue-specific protein profiles in normal and cancer tissues. *Mol. Cell. Proteomics* **2008**, *7* (5), 825−844.
(20) Meyer, M.; Kircher, M.; Gansauge, M. T.; Li, H.; Racimo, F.; Mallick, S.; Schraiber, J. G.; Jay, F.; Prufer, K.; de Filippo, C.; Sudmant, P. H.; Alkan, C.; Fu, Q.; Do, R.; Rohland, N.; Tandon, A.; Siebauer, M.; Green, R. E.; Bryc, K.; Briggs, A. W.; Stenzel, U.; Dabney, J.; Shendure, J.; Kitzman, J.; Hammer, M. F.; Shunkov, M. V.; Derevianko, A. P.; Patterson, N.; Andres, A. M.; Eichler, E. E.; Slatkin, M.; Reich, D.; Kelso, J.; Paabo, S. A High-Coverage Genome Sequence from an Archaic Denisovan Individual. *Science* **2012**, *338*, 222−226.
(21) Chen, Y.; Cunningham, F.; Rios, D.; McLaren, W. M.; Smith, J.; Pritchard, B.; Spudich, G. M.; Brent, S.; Kulesha, E.; Marin-Garcia, P.; Smedley, D.; Birney, E.; Flicek, P. Ensembl variation resources. *BMC Genomics* **2010**, *11*, 293.
(22) ENCODE Project Consortium. A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS Biol* **2011**, *9*, e1001046.
(23) Matys, V.; Kel-Margoulis, O. V.; Fricke, E.; Liebich, S.; Land, S.; Barre-Dirrie, A.; Reuter, I.; Chekmenev, D.; Krull, M.; Hornischer, K.; Voss, N.; Stegmaier, P.; Lewiki-Potapov, B.; Saxel, H.; Kel, A. E.; Wingender, E. TRANSFAC® and its module TRANSCompel®: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res.* **2006**, *34*, D108−D110.
(24) Lawrence, M.; Gentleman, R.; Carey, V. rtracklayer: an R package for interfacing with genome browsers. *Bioinformatics* **2009**, *25* (14), 1841−1842.
(25) Brawand, D.; Soumillon, M.; Necsulea, A.; Julien, P.; Csardi, G.; Harrigan, P.; Weier, M.; Liechti, A.; Aximu-Petri, A.; Kircher, M.; Albert, F. W.; Zeller, U.; Khaitovich, P.; Grutzner, F.; Bergmann, S.; Nielsen, R.; Paabo, S.; Kaessmann, H. The evolution of gene expression levels in mammalian organs. *Nature* **2011**, *478* (7369), 343−348.
(26) Anders, S.; Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **2010**, *11* (10), R106.
(27) Schwanhausser, B.; Busse, D.; Li, N.; Dittmar, G.; Schuchhardt, J.; Wolf, J.; Chen, W.; Selbach, M. Global quantification of mammalian gene expression control. *Nature* **2011**, *473* (7347), 337−342.
(28) Enard, W. Functional primate genomics–leveraging the medical potential. *J. Mol. Med. (Heidelberg, Ger.)* **2012**, *90* (5), 471−480.
(29) Zhuang, Y.; Soriano, P.; Weintraub, H. The helix-loop-helix gene E2A is required for B cell formation. *Cell* **1994**, *79* (5), 875−884.
(30) Dobreva, G.; Chahrour, M.; Dautzenberg, M.; Chirivella, L.; Kanzler, B.; Farinas, I.; Karsenty, G.; Grosschedl, R. SATB2 is a multifunctional determinant of craniofacial patterning and osteoblast differentiation. *Cell* **2006**, *125* (5), 971−986.
(31) Leoyklang, P.; Suphapeetiporn, K.; Siriwan, P.; Desudchit, T.; Chaowanapanja, P.; Gahl, W. A.; Shotelersuk, V. Heterozygous nonsense mutation SATB2 associated with cleft palate, osteoporosis, and cognitive defects. *Hum. Mutat.* **2007**, *28* (7), 732−738.
(32) Steidl, U.; Steidl, C.; Ebralidze, A.; Chapuy, B.; Han, H. J.; Will, B.; Rosenbauer, F.; Becker, A.; Wagner, K.; Koschmieder, S.; Kobayashi, S.; Costa, D. B.; Schulz, T.; O'Brien, K. B.; Verhaak, R. G.; Delwel, R.; Haase, D.; Trumper, L.; Krauter, J.; Kohwi-Shigematsu, T.; Griesinger, F.; Tenen, D. G. A distal single nucleotide polymorphism alters long-range regulation of the PU.1 gene in acute myeloid leukemia. *J. Clin. Invest.* **2007**, *117* (9), 2611−2620.
(33) Magnusson, K.; de Wit, M.; Brennan, D. J.; Johnson, L. B.; McGee, S. F.; Lundberg, E.; Naicker, K.; Klinger, R.; Kampf, C.; Asplund, A.; Wester, K.; Gry, M.; Bjartell, A.; Gallagher, W. M.; Rexhepaj, E.; Kilpinen, S.; Kallioniemi, O. P.; Belt, E.; Goos, J.; Meijer, G.; Birgisson, H.; Glimelius, B.; Borrebaeck, C. A.; Navani, S.; Uhlen, M.; O'Connor, D. P.; Jirstrom, K.; Ponten, F. SATB2 in combination

with cytokeratin 20 identifies over 95% of all colorectal carcinomas. *Am. J. Surg. Pathol.* **2011**, *35* (7), 937−948.
(34) Britanova, O.; Depew, M. J.; Schwark, M.; Thomas, B. L.; Miletich, I.; Sharpe, P.; Tarabykin, V. Satb2 haploinsufficiency phenocopies 2q32-q33 deletions, whereas loss suggests a fundamental role in the coordination of jaw development. *Am. J. Hum. Genet.* **2006**, *79* (4), 668−678.
(35) Wang, Y.; Li, Y.; Fan, X.; Zhang, Y.; Wu, J.; Zhao, Z. Early proliferation alteration and differential gene expression in human periodontal ligament cells subjected to cyclic tensile stress. *Arch Oral Biol.* **2011**, *56* (2), 177−186.