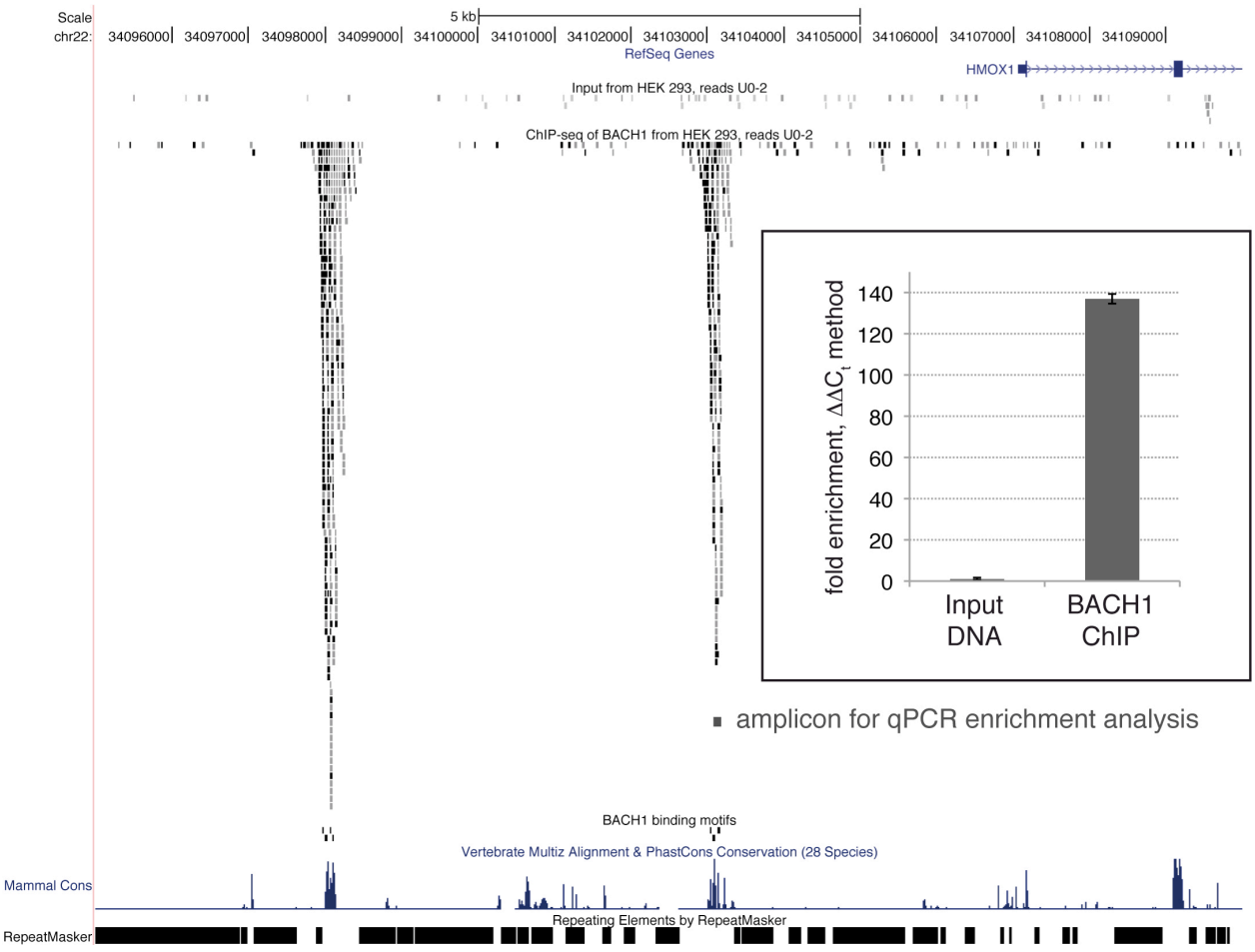
CONTENTS OF THE SUPPLEMENTAL INFORMATION

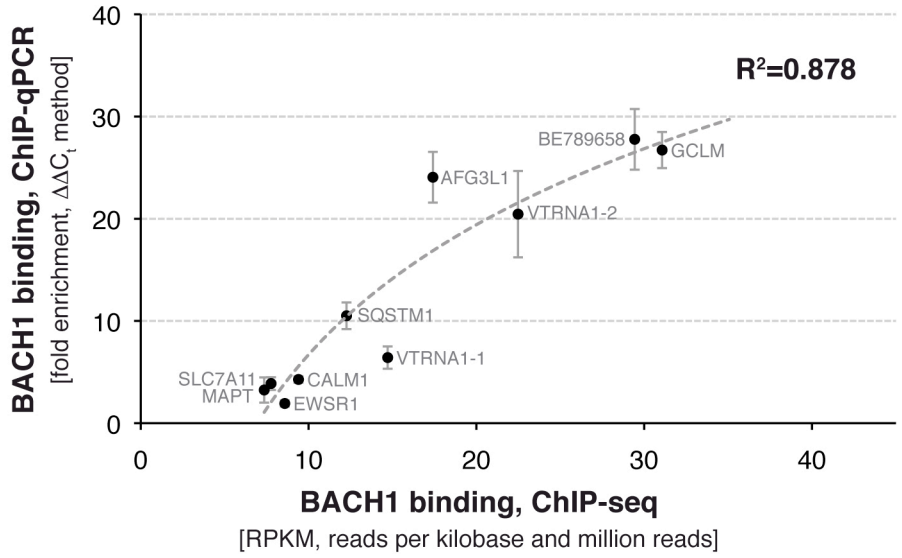
from Warnatz HJ et al. JBC 2011
"The BTB and CNC homology 1 (BACH1) target genes are involved in the oxidative stress response and in the control of the cell cycle"

SUPPLEMENTAL FIGURES

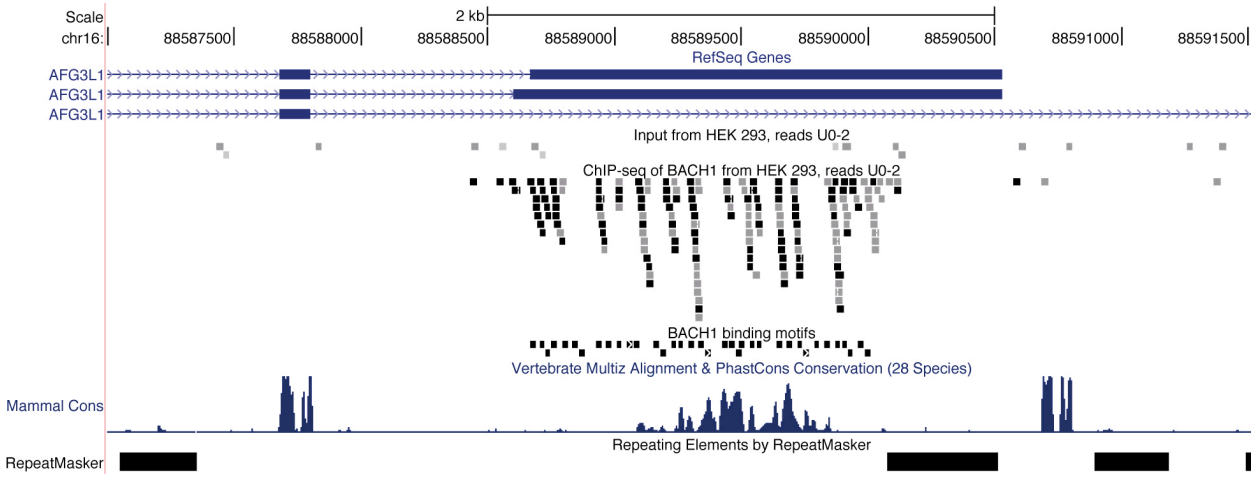
Supplemental Fig. 1. BACH1 ChIP-seq read enrichment in the distal HMOX1 promoter. A screenshot from the UCSC genome browser (human genome hg18 at chr22:34,095,001-34,110,000) shows the promoter region and start of the HMOX1 gene. ChIP-seq reads aligned to the forward strand are shown in black, reads aligned to the reverse strand are shown in grey. No read enrichment is visible in the control (input DNA sample), while the BACH1 ChIP sample shows strong read enrichments in two distal promoter regions. These conserved, non-repetitive binding regions at -4 kb and -9 kb upstream of the TSS were found to harbor three and four BACH1 binding motifs, respectively. The amplicon in the -4 kb region used for qPCR analysis is also shown in the screenshot. Inlay: Enrichment analysis of the indicated amplicon at -4 kb by ChIP-qPCR using the $\Delta\Delta C_t$ method showed 137-fold enrichment of the ChIP DNA compared to the input DNA. The enrichment was calculated from the ΔC_t of the HMOX1 -4 kb promoter region compared to the ΔC_t of a background region without BACH1 binding in the PSEN1 promoter.



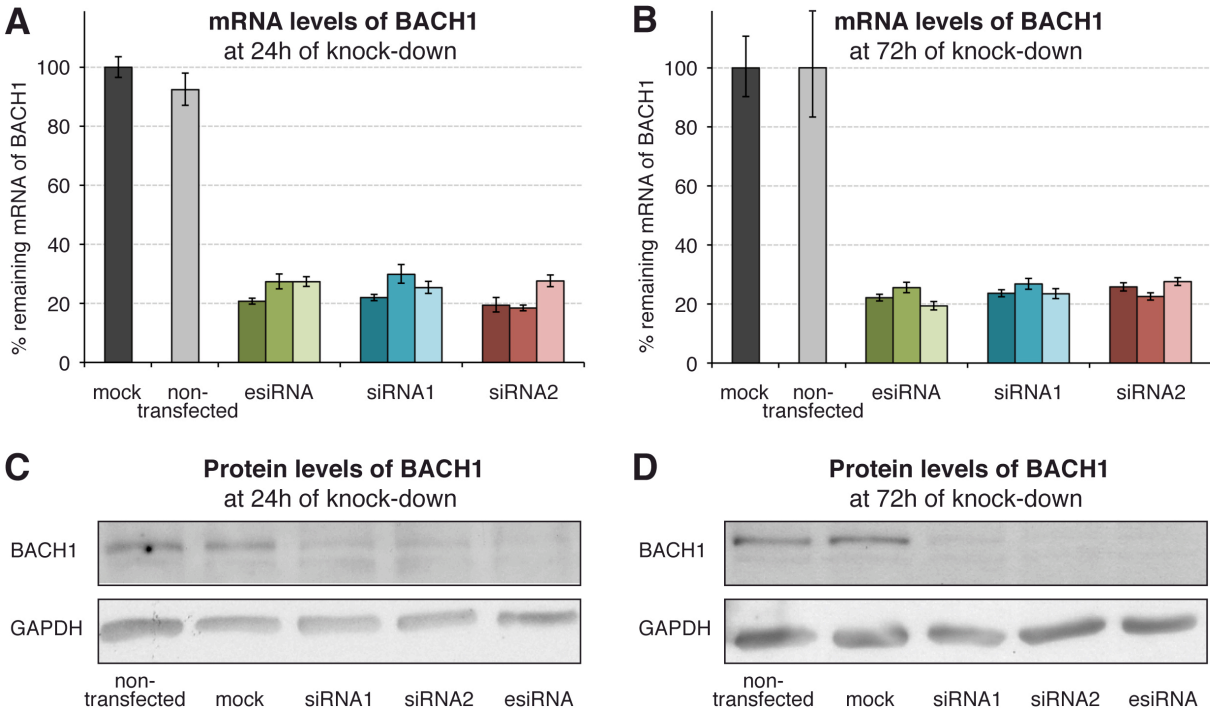
Supplemental Fig. 2. Validation of the ChIP-seq results by quantitative real-time PCR. The DNA from two independent ChIP samples from cells grown at different dates was used for qPCR analysis of ten genomic BACH1 target regions, ranging from strong to weak binding according to read density, measured in triplicates. A high correlation was observed (R^2=0.878) between RPKM values (reads per kilobase and million mapped reads) from ChIP-seq and the fold enrichment values ($\Delta\Delta C_t$ method) from ChIP-qPCR. Note that for the AFG3L1 gene, where multiple overlapping BACH1 binding peaks were present, the RPKM value of the qPCR amplicon was plotted instead of the RPKM value of one of the overlapping ChIP-seq peaks.



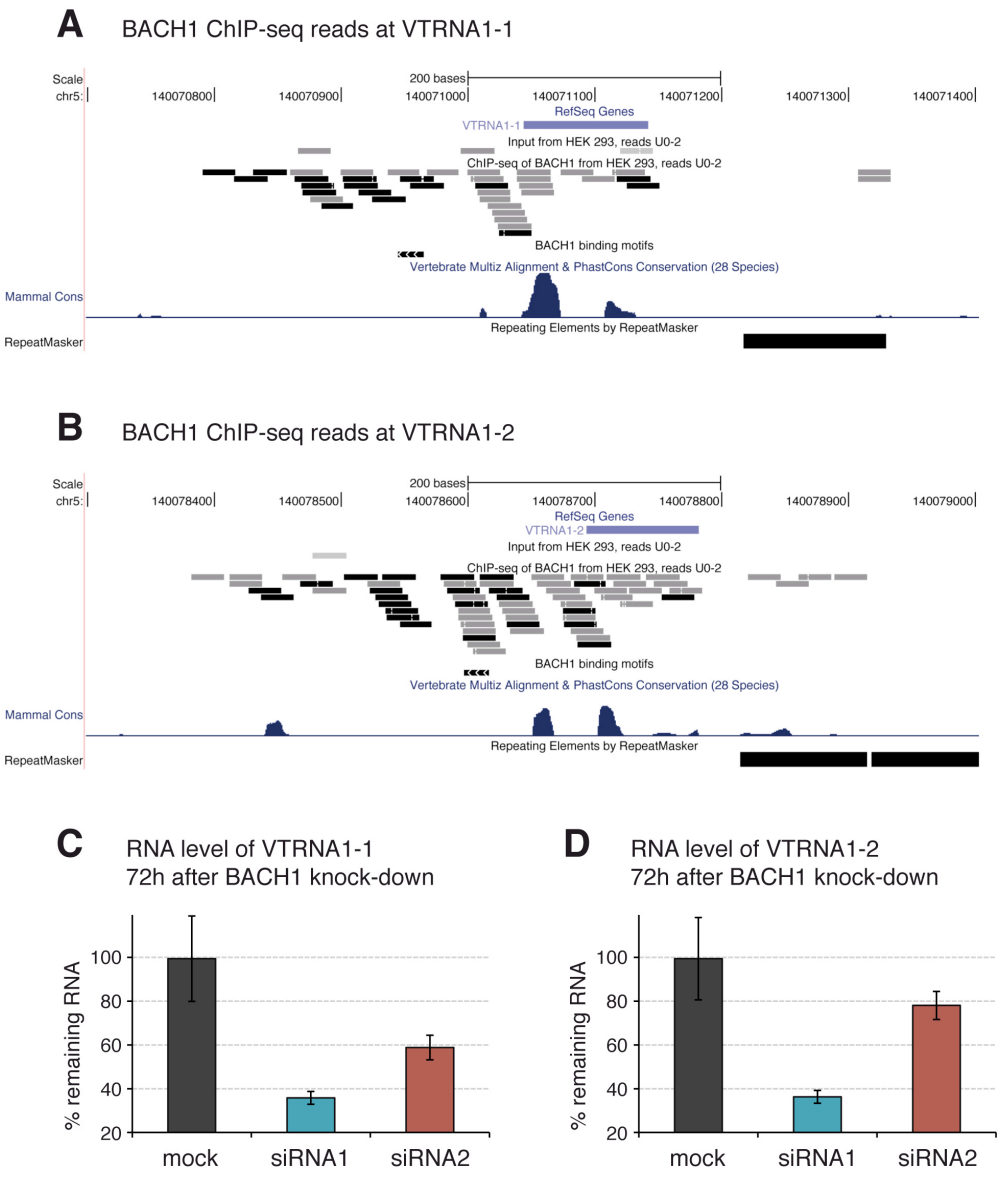
Supplemental Fig. 3. BACH1 binding at the non-coding RNA gene AFG3L1. The screenshot from the UCSC genome browser (human genome hg18 at chr16:88,587,000-88,591,500) shows the strong BACH1 binding at a large cluster of 36 BACH1 binding motifs in an alternative 3' exon of the AFG3L1 gene (RefSeq transcript accessions NR_003226 and NR_003227), which is located in an intron of another AFG3L1 splice variant (NR_003228). Eleven motif pairs showed a distance of 21 base pairs between motifs, pointing to BACH1 homodimerisation or heterodimerisation with small MAF proteins at the AFG3L1 gene. ChIP-seq reads aligned to the forward strand are shown in black, reads aligned to the reverse strand are shown in grey.



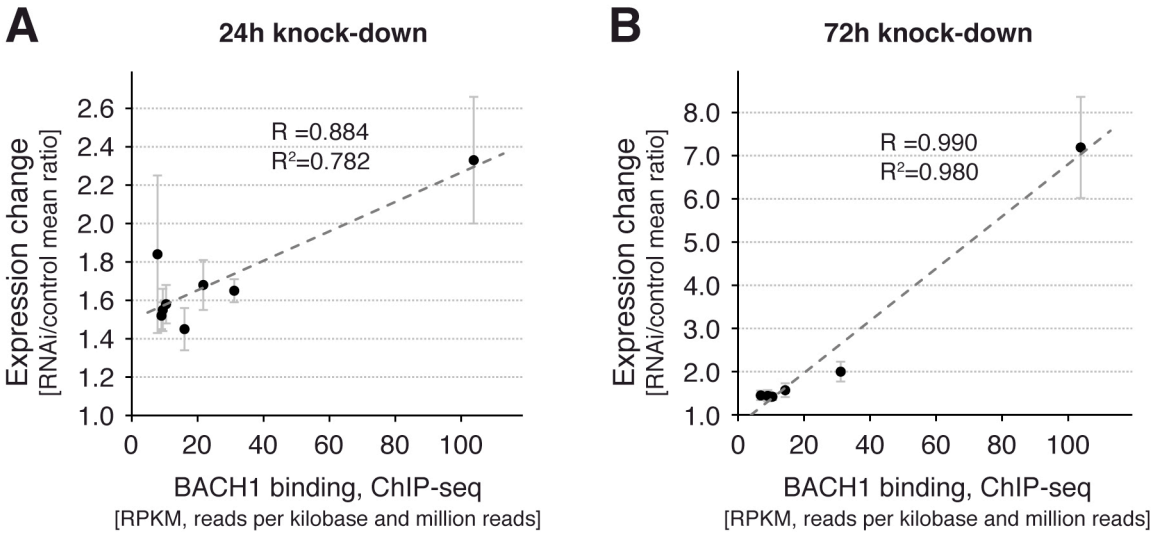
<u>Supplemental Fig. 4.</u> Efficiency of the BACH1 knock-down at the transcript and protein levels. *A*, at 24h post transfection, the amounts of BACH1 mRNA were reduced to 22-26% of the levels measured for the mock transfections, as calculated from the results of reverse transcription followed by quantitative real-time PCR in triplicates as fold-changes compared to beta-actin according to the $\Delta\Delta C_t$ method.. *B*, at 72h post transfection, the amounts of BACH1 mRNA were reduced to 22-25% of the levels measured for the mock transfections. *C*, at 24h post transfection, the amounts of BACH1 protein were reduced to <20% of the level measured for the mock transfection, as determined by Western blotting and immunodetection with anti-BACH1 antibody and anti-GAPDH antibody (loading control). *D*, at 72h post transfection, the amounts of BACH1 protein were reduced to <5% of the level measured for the mock transfection.



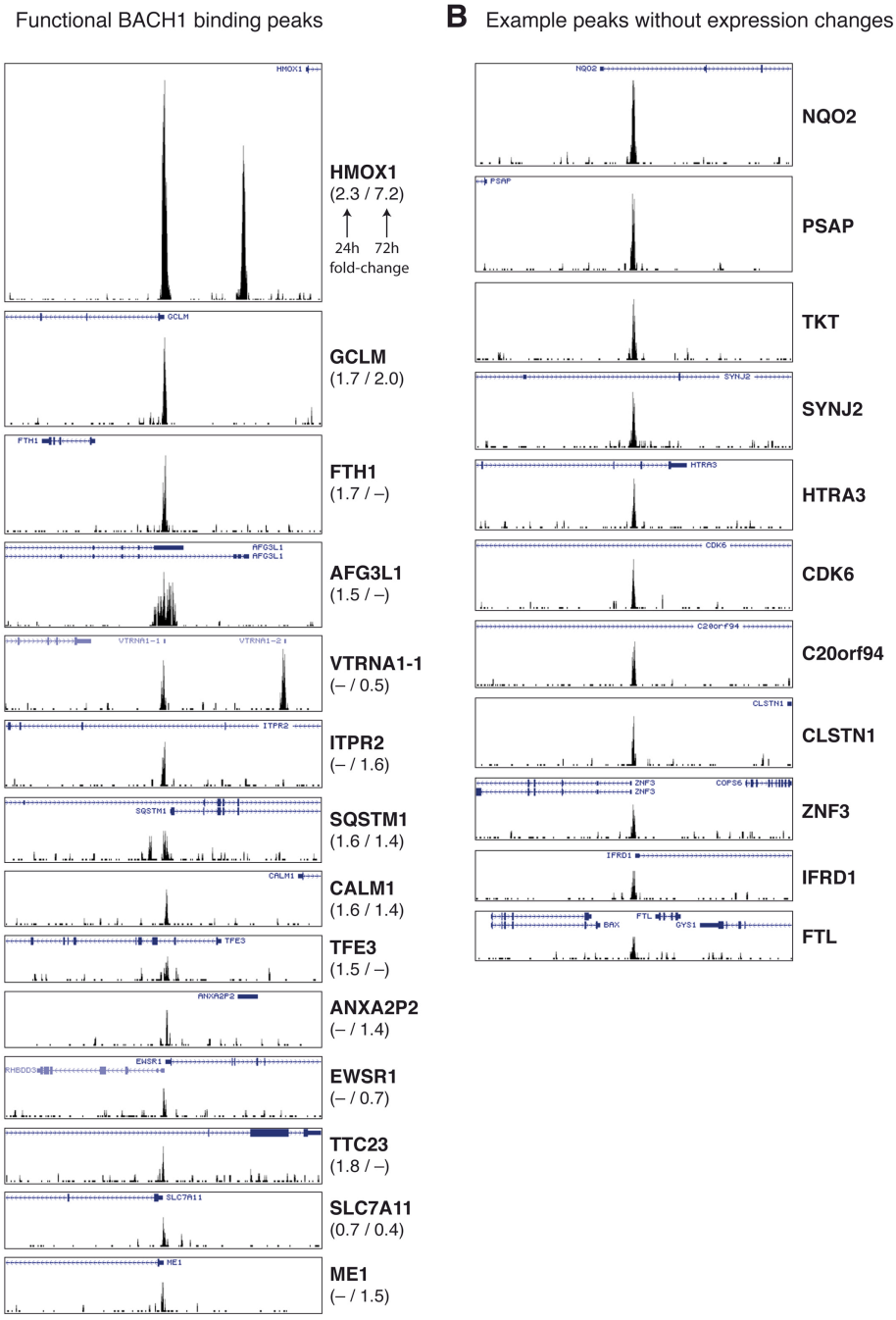
<u>Supplemental Fig. 5.</u> BACH1 binding at vault RNAs and expression changes after BACH1 knock-down. *A*, The screenshot from the UCSC genome browser (hg18 at chr5:140,070,700-140,071,400) shows the enrichment of BACH1 ChIP-seq reads at a BACH1 binding motif upstream of the VTRNA1-1 gene. ChIP-seq reads aligned to the forward strand are shown in black, reads aligned to the reverse strand are shown in grey. *B*, The screenshot from the UCSC genome browser (hg18 at chr5:140,078,300-140,079,000) shows the enrichment of BACH1 ChIP-seq reads at a BACH1 binding motif upstream of the VTRNA1-2 gene. *C*, 72h after transfection of BACH1 siRNAs, the expression of VTRNA1-1 was found significantly reduced to 36% and 59% by siRNA1 and siRNA2, respectively. *D*, 72h after transfection of BACH1 siRNAs, the expression of VTRNA1-2 was found significantly reduced to 36% by siRNA1, while the reduction to 78% by siRNA2 was not significant. The expression changes for the vault RNA transcripts were calculated from the results of reverse transcription followed by quantitative real-time PCR in triplicates as fold-changes compared to beta-actin according to the $\Delta\Delta C_t$ method.



<u>Supplemental Fig. 6.</u> Correlation between BACH1 binding and expression changes of direct target genes. *A*, a putative correlation (R^2=0.782) can be observed between BACH1 binding signals (RPKM values in untransfected cells) and the expression changes of BACH1 direct target genes 24h after transfection of BACH1 siRNAs. Note that the correlation coefficient is skewed by the high values for HMOX1. *B*, a putative correlation (R^2=0.980) can be observed between BACH1 binding signals (RPKM values in untransfected cells) and the expression changes of BACH1 direct target genes 72h after transfection of BACH1 siRNAs. Note that the correlation coefficient is skewed by the high values for HMOX1.



Supplemental Fig. 7. Comparison of functional BACH1 binding regions and binding regions without impact on gene expression in HEK 293 cells. Screen shots from the UCSC genome browser (20 kilobase regions in human genome hg18) showing significant BACH1 binding peaks. The scaling of the graphs is the same for all peaks. *A*, Pictures of all 14 binding regions that were found associated with expression changes of nearby target genes. The genes are sorted by decreasing BACH1 binding strength. The fold changes of expression at 24h and 72h after BACH1 knock-down are indicated in brackets below the gene symbols. *B*, Pictures of exemplary target genes where BACH1 binding was observed, but no significant gene expression changes were detected after BACH1 knockdown. The genes are sorted by decreasing binding strength.



SUPPLEMENTAL METHODS

Chromatin immunoprecipitation

The human embryonic kidney cell line HEK 293T/17 (CRL-11268 from ATCC) was cultured in DMEM low glucose (Invitrogen) supplemented with 100 U/ml Penicillin/G-Streptomycin (Invitrogen) and 10% heat-inactivated FBS (Biochrom) at 37°C and 5% CO2. Chromatin immunoprecipitation from HEK 293 was performed as described previously (1). A final concentration of 1% formaldehyde was added to 50 million adherent cells in five 175 cm^2 cell culture dishes to crosslink the proteins to the DNA for 10 min at room temperature. After cell lysis, the nuclei were collected and sonicated with a Branson Sonifier W250 to shear the chromatin to a final size of 200-500 bp. For immunoprecipitation, we utilized the well-characterized goat polyclonal C-20 antibody (Santa Cruz Biotechnology sc-14700X) recognizing the C-terminus of the BACH1 protein, which was used before by others for chromatin immunoprecipitation, immunofluorescence analysis and Western blotting (2,3). The sonicated chromatin was cleared by centrifugation for 10 min with 10,000 g and incubated with 10 µg goat anti-BACH1 antibody coupled to Protein G-magnetic beads (Invitrogen) overnight at 4°C. The control input DNA was chromatin that was reverse crosslinked, digested with RNase A and Proteinase K, phenol-chloroform extracted, and purified by ethanol precipitation with glycogen as carrier. The ChIP DNA was purified in parallel.

Library construction and sequencing

Sequencing libraries were prepared according to the manufacturer's instructions for the Illumina 1G Genome Analyzer, with some modifications. End-repair for the ChIP and control DNA was carried out with T4 DNA polymerase, Klenow DNA polymerase and T4 polynucleotide kinase (all from New England Biolabs) for 30 min at 20°C. Following DNA purification with the DNA Clean&Concentrator-5 kit (Zymo Research), addition of adenosine overhangs was performed with Klenow 3′–5′ exo minus (New England Biolabs) for 30 min at 37 °C. After DNA purification with the DNA Clean&Concentrator-5 kit, sequencing adapters (Genomic Adapter Oligo mix from Illumina) were ligated using Quick T4 DNA Ligase (New England Biolabs) for 15 min at 20°C. Following DNA purification with the DNA Clean&Concentrator-5 kit, the adapter-modified DNA was amplified by PCR with the genomic PCR primers 1.1 and 2.1 (Illumina) using the Phusion Master Mix with HF Buffer (New England Biolabs). The PCR program was Step 1: 98°C for 30 sec, Step 2: 98°C for 10 sec, Step 3: 65°C for 30 sec, Step 4: 72°C for 30 sec, Step 5: goto Step 2 (17 times), Step 6: 72°C for 5 min. After DNA purification with the DNA Clean&Concentrator-5 kit, the DNA libraries were seperated on 2% agarose TAE gels. The 150–250 bp fragments were excised from the gel on a Dark Reader (Claire Chemical Research) and purified with a Qiagen MinElute Gel Extraction Kit (Qiagen). The purified libraries were sequenced on an Illumina 1G Genome Analyzer.

Genomic alignment of reads and peak calling

Sequencing reads of 26 bases length were aligned to the human genome (UCSC hg18) using Eland from the Illumina Gerald module (v1.27) allowing up to two mismatches without insertions or deletions. We obtained 8,153,703 ChIP reads and 8,008,492 control reads with unique match to the genome. Redundant reads were removed for identification of candidate peak regions enriched in ChIP reads as compared to control reads using QuEST-2.4 (4) with a ChIP seeding fold enrichment of 30, a ChIP extension fold enrichment of 3 and a ChIP to background fold enrichment of 3. Artifactual peaks with an absolute peak shift (distance between peaks of forward and reverse reads) of less than 20 and a Bonferroni-corrected p-value for read enrichment larger than 10^{-60} were filtered out to obtain the final list of 84 peak regions significantly enriched in ChIP reads. To normalize and compare the signals from different peaks, the ChIP-seq reads were quantified in RPKM values (reads per kilobase and million aligned reads) as described before (5). For genes with more than one BACH1 binding peak, the total RPKM values were calculated as the sum of the RPKM values of the individual peaks.

Analysis of binding motifs, conservation scores and nearby genes

For DNA binding motif identification, the 300 bp genomic sequences surrounding each peak were extracted and the MEME algorithm (6) was applied with all default parameters to yield overrepresented binding motifs. The strongest log-odds matrix from the MEME output was compared to known motifs in the Transfac database using TOMTOM (7). This matrix was also used for MAST motif search (8) of all binding motifs in the peak sequences with a p-value less than 10^{-4}. Average conservation scores for peaks and DNA binding motifs were determined using Galaxy (9) with phastCons on 17-species multiz alignment of UCSC hg18. Conservation plots were obtained from the Cis-regulatory Element Annotation System CEAS (10). Nearby genes within 15 kilobases of the peak regions were annotated using CisGenome v1.2 (11) using the UCSC refFlat table for the human genome (UCSC hg18).

Inhibition of BACH1 expression by RNA interference

To silence the endogenous expression of BACH1 in HEK 293 cells, we performed RNA interference experiments using three independent types of silencing molecules against BACH1: one unmodified synthetic small interfering RNA ("siRNA1") from Qiagen (product no. SI00309876), one chemically modified synthetic small interfering RNA ("siRNA2") from Invitrogen (product no. HSS100910), and one independent high-complexity pool of 20-30 bp siRNA-like molecules ("esiRNA") prepared by RNase III digestion of long dsRNA generated by *in vitro* transcription from T7-linked BACH1 RT-PCR products (chr21:29637334-29637895 in UCSC hg18) of cDNA from HEK 293 (12). The *E. coli* RNase III clone was kindly provided by Dr. Frank Buchholz. For transfection, ca. 18,000 cells/cm² were seeded in 12-well plates together with esiRNA/HiPerFect complexes (200 ng esiRNA/6 µl HiPerFect per well) or siRNA/HiPerFect complexes (300 ng siRNA/6 µl HiPerFect per well) according to the HiPerFect fast-forward protocol (Qiagen). For the mock transfections, cells were treated with HiPerFect reagent only. We performed each knock-down transfection in triplicates and each mock-transfection and non-transfection in quadruplicates.

RNA extraction, reverse transcription and hybridization on microarrays

Total RNA was extracted from cultured cells at 24h and 72h post transfection using the RNeasy Mini Kit (Qiagen) following the manufacturers instructions. All RNA samples were treated on-column with RNase-free DNase I, quantified by UV spectrophotometry and controlled for integrity by gel electrophoresis and capillary electrophoresis using the Agilent 2100 bioanalyzer. Reverse transcription reactions were performed with random hexamer primers and SuperScript II reverse transcriptase (Invitrogen). Complementary DNA for quantitative real-time PCR analysis was prepared from 1 µg of total RNA of each sample in 20 µl reactions and diluted to 12.5 ng/µl equivalent of total RNA. For hybridizations on microarrays, 1 µg of DNA-free total RNA from each sample was used to synthetize biotinylated cRNA using the GeneChip Expression 3' Amplification One-Cycle Target Labeling and Control Reagents kit from Affymetrix (P/N 900493). Following integrity control using an Agilent 2100 bioanalyzer, the cRNA was fragmented and hybridized to the Affymetrix GeneChip Human Genome U133 Plus 2.0 (HG-U133Plus2). The arrays were washed, stained, and scanned by the German Resource Center for Genome Research (RZPD) following recommended protocols from Affymetrix.

Analysis of gene expression changes after BACH1 knock-down

The knock-down efficiency of each silencing molecule was measured 24h and 72h after transfection, both at the BACH1 mRNA level (by quantitative real-time RT-PCR, see below) and at the protein level by Western blotting and immunodetection using goat anti-BACH1 (Santa Cruz Biotechnology sc-14700X). The genome-wide expression changes from the BACH1 knock-downs were analyzed in triplicates, together with four controls (mock transfections), using the results from the Affymetrix U133Plus2 array hybridizations. The array probe intensities from the knock-down samples were normalized together with those from the control samples using GCRMA from the R/Bioconductor package (13) to give p-values from a Student's t-test as described previously (14). The expression ratios for each probe were calculated as average of the ratios of the treated samples divided by the average of the ratios of the control samples.

Probes with significant expression changes were defined by a p-value<0.05 and expression ratio>1.3 (up-regulation threshold) or ratio<0.75 (down-regulation threshold). Genes with significant expression changes were defined by significant changes of the same probes among at least two out of three different RNAi experiments. The probability to find 13 or more out of 59 BACH1 target genes with expression changes after BACH1 knock-down was calculated with the hypergeometric test, given a set of 1,570 changed genes among 19,511 genes represented on the Affymetrix U133Plus2 array. Gene ontology enrichment analysis among up-regulated genes was performed using the DAVID Bioinformatics Resources 6.7 (15) in the GO term "biological process" at level 5 with an EASE score theshold of 0.05 and a minimum count of two genes in enriched categories.

Quantitative real-time PCR analysis

The ChIP enrichment as compared to input DNA was calculated from quantitative real-time PCR results as fold-enrichment according to the ΔΔCt method (16). PCR was performed in triplicates using SYBR Green (Applied Biosystems) with 20 pg|µl ChIP or input DNA and a final primer concentration of 750 nM. Detection primers for the known target region were HMOX1fwd 5'-GAA GGC GGA TTT TGC TAG ATT T-3' and rev 5'-CTC CTG CCT ACC ATT AAA GCT G-3', for the background region PSEN1fwd 5'-GAA ATG ACG ACA ACG GTG AG-3' and rev 5'-CTC AGG TTC CTT CCA GAC CA-3'. Primer sequences for independent qPCR validation of other genomic target regions were GCLMfwd 5'-CGG GAG AGC TGA TTC CAA AC-3' and rev 5'- GAA GCA CTT TCT CGG CTA CGA-3', BE789658fwd 5'-CTG ACG GGG GAA ATC CAC T-3' and rev 5'-GGC GGA TCT CTG CTG ACT C-3', VTRNA1-2fwd 5'-GGG AAG GCT GTG TCC TTG TC-3' and rev 5'-GCA ACC AGG ACT GTC CAA CA-3', AFG3L1fwd 5'-GGC TTA GTA CTG CCC CTC AG-3' and rev 5'-GCT GTG TCA TCG CTG CTA-3', VTRNA1-1fwd 5'-TCC CCA GAT GGA CAA CTC CT-3' and rev 5'-TGG TGA GAA AGA CCT ACG TCA CA-3', SQSTM1fwd 5'-CCT GAT ATG GGG GCT GTG TC-3' and rev 5'-GCA CCT GGG ATC AGG GTA CT-3', CALM1fwd 5'-AGG GAA GAG CTG GAG CAG TG-3' and rev 5'-CTC CAC CAG TCC CAT GCA AT-3', EWSR1fwd 5'-AAT CCA TTC CGC GCA CAC-3' and rev 5'-CCT GCA GGG AGA CGG AGA T-3', SLC7A11fwd 5'-CAG GTT TGC ATC AGC CAC AT-3' and rev 5'-TGA GCA ACA AGC TCC TCC TG-3', MAPTfwd 5'-GCT TTC CCC AGA CCA GAA CC-3' and rev 5'-GGG CAG CCA AGG AAG GTC-3', IL4(control)fwd 5'-CAA GAT GCC ACC TGT ACT TGG A-3' and rev 5'-CCA CAG GTG TCC GAA TTT GTT-3'. The knock-down efficiency for the BACH1 mRNA and expression changes of the vault RNAs VTRNA1-1 and VTRNA1-2 were calculated from quantitative real-time PCR results as fold-changes compared to beta-actin (ACTB) according to the ΔΔCt method (16). PCR was performed in triplicates using SYBR Green (Applied Biosystems) with a cDNA concentration of 12.5 ng|µl RNA equivalent from reverse transcription and a final primer concentration of 750 nM. Detection primers for PCR were BACH1RTfwd 5'-GTC CTT GTT GAA AAT GCA CAA G-3' and rev 5'-AGG AAA TTC CCT TAT GGT AAA CTA-3' (detect both BACH1 mRNA isoforms), VTRNA1-1RTfwd 5'-GCT TTA GCT CAG CGG TTA CTT CG-3' and rev 5'-GGG TCT CGA ACA ACC CAG ACA-3', VTRNA1-2RTfwd 5'-GCT GGC TTT AGC TCA GCG GTT A-3' and rev 5'-GGG TCT CGA ACC ACC CAG AG-3' and ACTB(control)RTfwd 5'-TCA AGA TCA TTG CTC CTC CTG AG-3' and rev 5'-ACA TCT GCT GGA AGG TGG ACA-3'.

Correlation between ChIP-seq reads and expression ratios

Mean expression ratios from the different RNAi experiments were calculated as arithmetic mean of the individual ratios. The standard errors for the mean ratios were calculated as the root of the mean squared standard errors of the individual RNAi experiments. To assess the correlation between RPKM values and RNAi expression ratios, the Pearson correlation was calculated, which is not optimally suited because the correlation is expected to be non-linear and the relatively high values for HMOX1 skew the correlation coefficient. Therefore, also Spearman's rank correlation coefficient was used as more robust non-parametric measure of statistical dependence between the two variables to assess how well a relationship can be described using a monotonic function. The expression ratios for BACH1 were not included in the tests.

Transcription factor affinity prediction in promoter sequences

Up-regulation after BACH1 knock-down was observed for 883 genes (760 genes after 24h and 174 genes after 72h, with 51 genes common to both time points), and down-regulation was observed for 728 genes (147 genes after 24h and 663 genes after 72h, with 82 genes common to both time points). The RefSeq identifiers for the associated transcripts were retrieved from the gene information file provided by the supplier of the Affymetrix array HG-U133 Plus 2.0. Thus, we obtained 829 RefSeq identifiers with mappable TSS for the up-regulated genes and 687 RefSeq identifiers with mappable TSS for the down-regulated genes. The remaining TSSs could not be retrieved because the RefSeq identifiers had been removed from the NCBI's Reference Sequence database during genomic annotation updates (6.1% of transcripts from up-regulated genes and 5.6% of transcripts from down-regulated genes). Using the obtained 829 TSS coordinates of up-regulated genes and 687 TSS coordinates of down-regulated genes, we retrieved promoter sequences of 2 kb length, stretching 1.8 kb upstream to 200 bp downstream of the respective TSS. The two retrieved sequence sets were used for transcription factor affinity prediction, taking into account all possible binding sites within the sequences (17). Promoter sequences were scanned for transcription factors that may co-regulate the set of sequences of using 554 TF binding matrices in TransFac (version 12.1) and a human promoter-based background model. P-values for the individual sequences were combined by Fisher's method and multiple test-corrected according to Benjamini-Hochberg, giving a natural ranking of the transcription factors that have the most enriched binding within the whole sequence set. The top-ranking factors described in the Results section met the additional thresholds of a combined and corrected $p<10^{-29}$ and a mean affinity score across the whole sequence sets of at least 0.01 (for up-regulated genes) or 0.005 (for down-regulated genes).

SUPPLEMENTAL REFERENCES

1. Schmidt, D., Wilson, M. D., Spyrou, C., Brown, G. D., Hadfield, J., and Odom, D. T. (2009) *Methods* **48**, 240-248
2. Dhakshinamoorthy, S., Jain, A. K., Bloom, D. A., and Jaiswal, A. K. (2005) *J. Biol. Chem.* **280**, 16891-16900
3. Sakamoto, K., Iwasaki, K., Sugiyama, H., and Tsuji, Y. (2009) *Mol. Biol. Cell* **20**, 1606-1617
4. Valouev, A., Johnson, D. S., Sundquist, A., Medina, C., Anton, E., Batzoglou, S., Myers, R. M., and Sidow, A. (2008) *Nat. Methods* **5**, 829-834
5. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., and Wold, B. (2008) *Nat. Methods* **5**, 621-628
6. Bailey, T. L., and Elkan, C. (1994) *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **2**, 28-36
7. Gupta, S., Stamatoyannopoulos, J. A., Bailey, T. L., and Noble, W. S. (2007) *Genome Biol.* **8**, R24
8. Bailey, T. L., and Gribskov, M. (1998) *Bioinformatics* **14**, 48-54
9. Taylor, J., Schenck, I., Blankenberg, D., and Nekrutenko, A. (2007) *Curr Protoc Bioinformatics* **Chapter 10**, Unit 10 15
10. Ji, X., Li, W., Song, J., Wei, L., and Liu, X. S. (2006) *Nucleic Acids Res.* **34**, W551-554
11. Ji, H., Jiang, H., Ma, W., Johnson, D. S., Myers, R. M., and Wong, W. H. (2008) *Nat. Biotechnol.* **26**, 1293-1300
12. Kittler, R., Surendranath, V., Heninger, A. K., Slabicki, M., Theis, M., Putz, G., Franke, K., Caldarelli, A., Grabner, H., Kozak, K., Wagner, J., Rees, E., Korn, B., Frenzel, C., Sachse, C., Sonnichsen, B., Guo, J., Schelter, J., Burchard, J., Linsley, P. S., Jackson, A. L., Habermann, B., and Buchholz, F. (2007) *Nat. Methods* **4**, 337-344
13. Gentleman, R. C., Carey, V. J., Bates, D. M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., Hornik, K., Hothorn, T., Huber, W., Iacus, S., Irizarry, R., Leisch, F., Li, C., Maechler, M., Rossini, A. J., Sawitzki, G., Smith, C., Smyth, G., Tierney, L., Yang, J. Y., and Zhang, J. (2004) *Genome Biol.* **5**, R80
14. Herwig, R., Aanstad, P., Clark, M., and Lehrach, H. (2001) *Nucleic Acids Res.* **29**, E117
15. Dennis, G., Jr., Sherman, B. T., Hosack, D. A., Yang, J., Gao, W., Lane, H. C., and Lempicki, R. A. (2003) *Genome Biol.* **4**, P3

16. Livak, K. J., and Schmittgen, T. D. (2001) *Methods* **25**, 402-408
17. Roider, H. G., Kanhere, A., Manke, T., and Vingron, M. (2007) *Bioinformatics* **23**, 134-141