



(51) International Patent Classification:

C07K 14/415 (2006.01) *C12N 15/82* (2006.01)
C12N 15/62 (2006.01) *C12N 15/10* (2006.01)

(21) International Application Number:

PCT/EP2013/065695

(22) International Filing Date:

25 July 2013 (25.07.2013)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

12179185.9 3 August 2012 (03.08.2012) EP

(71) Applicants: **VIB VZW** [BE/BE]; Rijvisschestraat 120, B-9052 Gent (BE). **UNIVERSITEIT GENT** [BE/BE]; Sint-Pietersnieuwstraat 25, B-9000 Gent (BE).

(72) Inventors: **DE JAEGER, Geert**; Azaleastraat 51, B-9940 Evergem (BE). **VERKEST, Aurine**; K.L. Dierickxstraat 10, B-9000 Gent (BE). **INZÉ, Dirk, Gustaaf**; Kortenbosdries 18, B-9310 Moorsel - Aalst (BE). **VANDEPOELE, Klaas**; Steenstraat 5B2, B-9630 Zwalm (BE). **ABEEL, Thomas**; Oostweg 2, B-8700 Tielt (BE).

(74) Common Representative: **VIB VZW**; Rijvisschestraat 120, B-9052 Gent (BE).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

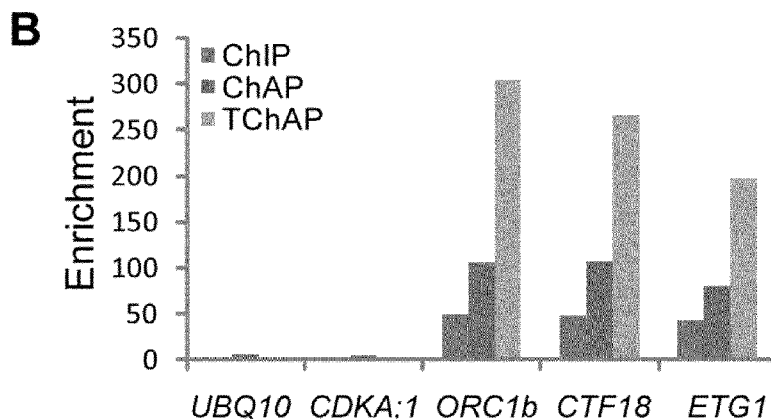
(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

- with international search report (Art. 21(3))
- with sequence listing part of description (Rule 5.2(a))

(54) Title: TANDEM CHROMATIN AFFINITY PURIFICATION IN PLANT CELL SUSPENSION CULTURES FOR IDENTIFICATION OF TRANSCRIPTION FACTOR TARGET GENES

Figure 2



(57) Abstract: The present invention relates to the identifications of target genes of transcription factors in plants. More specifically, it relates to a combination of chromatin immunoprecipitation with Tandem affinity purification, applied to plant cells.

Tandem chromatin affinity purification in plant cell suspension cultures for identification of transcription factor target genes

The present invention relates to the identification of target genes of transcription factors in plants. More specifically, it relates to a combination of chromatin immunoprecipitation with Tandem affinity purification, applied to plant cells.

Regulation of gene expression plays an important role in a variety of biological processes such as growth, development and responses to environmental stimuli. Many regulatory genes encode transcription factors (TFs), which modulate gene expression by binding to regulatory sequences of their target genes. Which genes are directly controlled by these TFs, and the molecular mechanisms of target gene recognition *in vivo* are, particularly in plants, however still largely unexplored.

A common used tool to address these questions is chromatin immunoprecipitation (ChIP) (Kuo and Allis 1999). In a classical ChIP experiment, DNA fragments associated with a specific protein are enriched. DNA binding protein complexes are reversibly crosslinked with formaldehyde, the chromatin is fragmented, and the DNA fraction that interacts with the TF of interest is isolated by immunoprecipitation with a specific antibody. Finally, DNA sequences associated with the precipitated protein can be identified by hybridization to tiling arrays (ChIP-chip) or by direct high-throughput sequencing (ChIP-seq) (Kim and Ren 2006; Park 2009).

In spite of its power, conventional ChIP has experimental boundaries. Its main shortcoming, especially in the case of genome-wide applications, is the overall inefficiency of ChIP enrichment. This drawback, which is a consequence of crosslinking and compromises the identification of low-abundant TF-DNA interactions, necessitates the requirement for large cell numbers and the need for high-quality antibodies. Reports of genome-wide ChIP studies of plant TFs are, in comparison to other eukaryotic systems, still lagging behind. Although plant-specific ChIP protocols have been successfully developed (Bowler et al. 2004; Gendrel et al. 2005; Saleh et al. 2008; Kaufmann et al. 2010), plant features, such as rigid cell walls, large vacuoles, chloroplasts, and the paucity of nuclei in some tissues, combined with TF tissue and target specificities, all challenge TF-DNA enrichment.

Surprisingly we found that an alternative ChIP protocol relying on *Arabidopsis thaliana* cell suspension cultures and TF tandem affinity tagging showed a superior performance in isolation of the TF target genes. The development of a tandem affinity purification (TAP) platform in plant cell suspension cultures to study protein interactomes was described previously (Van Leene et al. 2007; Van Leene et al. 2008). Through combining the TAP approach with ChIP protocol features, a tandem chromatin affinity purification (TChAP) method was established for assaying *in vivo* TF-DNA binding. TChAP circumvents the necessity of

highly specific antibodies and allows the identification of more and less prevalent TF-DNA interactions enhancing both data quality and quantity of TF location studies.

A first aspect of the invention is a method for isolating target genes for transcription factors in plants, comprising (1) transforming a plant cell with an expression vector comprising a transcription factor fused to at least two tags (2) expressing said transcription factor in the plant cell (3) crosslinking the proteins with the DNA (4) extracting the proteins and chromatin from the cell (5) fragmenting the chromatin (6) purifying the chromatin-protein complex using affinity binding towards the first tag and (7) purifying the fraction comprising the first tag using affinity binding towards the second tag. A "tag" as used here is a peptide or polypeptide that can bind another compound in a specific way; said compound can be any inorganic or organic chemical compound, including but not limited to a metal or another peptide or polypeptide. Tags are known to the person skilled in the art, and include but are not limited to a polyhistidine tag, FLAG tag, HA tag, MYC tag, EPAE tag, GST tag, and biotin tag. Preferably at least one of the tags is selected from the group consisting of a polyhistidine tag and a biotine tag. Even more preferably, one tag is a polyhistidine tag and one tag is a biotine tag. Most preferably the tags are comprised in a polyhistidine- biotin-polyhistidine construct. The fusion of the tags to the transcription factor may be in any way known to the person skilled in the art, at the carboxyterminal end, at the aminoterminal end, or internally, as long as the transcription factor is still capable of binding its target DNA. There may be more than two tags, and each of the tags may be used for purification. Tag may be fused in tandem, or may be separated by non-tag sequences.

An "affinity binding" as used here is any kind of specific binding, including, but not limited to antigen-antibody binding, or protein-chemical compound binding. Preferably, the affinity binding is carried out by using a binding compound linked to a carrier, such as, but not limited to a nitrocellulose membrane, a resin, a sepharose bead or an agarose gel. Preferably, said affinity binding is selected from the group consisting of a nickel chelating carrier and a streptavidine carrier. Even more preferably said affinity binding is selected from the group consisting of a Ni-NTA resin and Streptavidin sepharose beads. Most preferably, one affinity binding is carried out with a Ni-NTA resin and the other with Streptavidin sepharose beads.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1. E2Fa-HBH expression and biotinylation analysis. (A) E2Fa-HBH protein expression analysis. Immunoblot analysis of wild-type (PSB-D) and *E2Fa-HBH* overexpressing cell suspension cultures using anti-E2Fa, anti-His antibodies and a Strep-HRP conjugate. The arrowhead indicates the endogenous E2Fa and the asterisk the ectopic E2Fa-HBH detected with the anti-E2Fa antibody. (B) E2Fa-HBH *in vivo* biotinylation efficiency. Western blotting with an anti-His antibody to monitor binding of E2Fa-HBH to Streptavidin Sepharose beads.

Information on the percentage of input and the estimation of the bead bound amount is indicated.

Figure 2. Comparison of different chromatin isolation methods. (A) E2Fa-HBH protein purification efficiency. Immunoblot analysis using an anti-His antibody to evaluate the ChIP, ChAP and TChAP E2Fa-HBH protein purification yields. Information on the percentage of input and the estimation of the purification efficiency is given. (B) ChIP, ChAP and TChAP DNA qPCR analysis. Association of E2Fa with the proximal promoters of negative control genes *CDKA;1* and *UBQ10* and of the well established E2Fa target genes, *ORC1b*, *CTF18* and *ETG1*, was quantified by qPCR. The enrichment of E2Fa binding was normalized against the input DNA.

Figure 3. ChIP, ChAP and TChAP E2Fa majority genes. (A) Venn diagrams showing overlap among all, the E2Fa regulated and the E2F motif majority genes identified through ChIP-, ChAP- and TChAP-seq. (B) Examples of known E2Fa binding sites identified in the ChIP-, ChAP- and TChAP-seq data. Visual representation is organized in four tracks. Gene structure is indicated with cyan rectangles. The normalized coverage for each of the strategies is displayed as a bar chart around a central axis with coverage on the forward and reverse shown in green (up) and blue (down), respectively. Total coverage is indicated in yellow. The coverage scale is the same for all three protocols and each track has been normalized for total coverage in that protocol, allowing direct comparison. (C) Frequency of E2Fa regulated or E2F motif genes among the ChIP, ChAP and TChAP identified majority peaks. Peaks are sorted based on descending MACS confidence peak score. The grey lines depict the random occurrence of an E2Fa regulated gene in the *Arabidopsis* genome and the expected frequency of E2F motif genes based on randomized peak sequences, respectively.

Figure 4. E2Fa TChAP-seq data analysis. (A) Distribution of E2Fa binding sites in the *Arabidopsis* genome. (B) Venn diagram showing the overlap of the TChAP E2Fa bound genes with E2Fa regulated genes and genes containing an E2F motif in their 1-kb promoter. (C) Overlap of TChAP E2Fa genes with RBR (Gutzat et al. 2011) and E2Fa (Naouar et al. 2009) regulated genes. (D) Enrichment analysis of the TChAP data for DNA repair, cell cycle, cell cycle periodic, G1/S-, and G2/M-phase expressed genes. Shown are negative $\log_{10}$ values of the likelihood to obtain an observed fraction of specific genes by change. The vertical line indicates p-value 0.002 (Bonferroni corrected p 0.05).

Figure 5. Comparison of ChIP, ChAP and TChAP TOP200 genes. Venn diagrams showing overlap among all, the E2Fa regulated and the E2F motif TOP200 genes identified by ChIP-, ChAP- and TChAP-seq.

- 5 **Figure 6.** Validation of unique TChAP majority genes. Binding of E2Fa to twelve TChAP unique identified binding regions, having different confidence scores, was checked by quantitative TChAP-PCR. Besides known E2Fa target genes (right) also new identified E2Fa bound genes (left) were selected. Two negative controls were included (CDKA;1 and UBQ10).
- 10 **Figure 7.** Validation of new E2Fa targets. Expression levels of TChAP identified potential new E2Fa target genes were examined by qRT-PCR in both wild-type (Col-0) and E2Fa-DPa overexpressing seedlings. The dots indicate genes validated in ChIP and/or ChAP.

EXAMPLES

15 **Materials and methods to the examples**

Constructs, plant material, transformation and growth conditions

- The HBH tag sequence was amplified by PCR using the pFA6HBH-KanMX6 plasmid as template (Tagwerker et al. 2006) and cloned into the pDONRP2RP3 ENTRY vector by a BP recombination reaction according to the manufacturer's instructions (Invitrogen). The
- 20 *P35S:E2Fa-HBH*-containing plant transformation vector was obtained by a Multisite Gateway LR reaction between pEntryL4R1-P35S, pEntryL1L2-E2Fa, and pEntryR2L3-HBH and the destination vector pK7GW43D (Van Leene et al. 2007). The construct was transferred into the *Agrobacterium tumefaciens* C58C1Rif^R strain harboring plasmid pMP90 for transformation of *Arabidopsis* cell suspension cultures. Maintenance and stable transformation of *Arabidopsis*
- 25 *thaliana* PSB-D cell suspension cultures was done according to Van Leene et al. (2007). *Arabidopsis thaliana* (L.) Heyhn. (ecotype Columbia-0) plants were grown under long-day conditions (16 h of light and 8 h of darkness) at 22°C on 0.5x Murashige and Skoog (MS) agar plates (Valvekens et al. 1988). The *E2Fa-DPa* overexpressing plants have been described previously (De Veylder et al. 2002).

30

Protein extraction, SDS-PAGE and immunoblotting

- Two-day-old PSB-D cell suspension cultures were harvested, used immediately or snap-frozen in liquid nitrogen, and stored at -70°C. Proteins were extracted after grinding in liquid nitrogen in homogenization buffer (25 mM Tris-Cl, pH 7.6, 75 mM NaCl, 15 mM MgCl₂, 15 mM EGTA,
- 35 15 mM p-nitrophenylphosphate, 60 mM β-glycerophosphate, 1 mM DTT, 0.1% Nonidet P-40, 0.1 mM Na₃VO₄, 1 mM NaF, and protease inhibitor cocktail P9599 (Sigma-Aldrich)).

Proteins were separated by 10% SDS-PAGE and blotted onto Immobilon-P membranes (Millipore, Bedford, MA). Filters were blocked in 3% (v/v) milk powder in 25 mM Tris-Cl (pH 8), 150 mM NaCl, 0.05% Tween 20 for at least 1 h at room temperature and incubated overnight at 4°C with E2Fa (1/2500) (Takahashi et al. 2008), His (1/2000)(Qiagen), or Streptavidin-HRP (1/4000) (Amersham Biosciences) antibody in blocking buffer. Antigen-antibody complexes were detected with horseradish peroxidase-conjugated IgG diluted 1/10000 (Amersham Biosciences) with a chemiluminescence system (Perkin Elmer, Norwalk, CT). Purification efficiency or bait recovery was estimated from western blot signal intensities using the Biorad Image Lab 3.0 software.

RNA isolation, cDNA preparation and qRT-PCR

RNA was extracted using the RNeasy Plant Mini Kit (Qiagen) and DNase I treated (Promega) prior to cDNA synthesis. cDNA was prepared from DNase I treated total RNA with the iScript cDNA Synthesis Kit (Biorad) according to the manufacturer's instructions. Relative transcript abundance of selected genes was determined using the Roche LightCycler 480 system and the LC480 SYBR Green I Master kit (Roche Diagnostics). Measurements were taken for two biological and three technical repeats. The amplification data were analyzed using the second derivative maximum method, and resulting Cp values were converted into relative expression values using the comparative Ct method.

Chromatin immunoprecipitation (ChIP), chromatin affinity purification (ChAP) and tandem chromatin affinity purification (TChAP)

Two-day-old exponentially growing cell cultures were treated with 0.75% formaldehyde for 10 min. Crosslinking was stopped by addition of 0.25 M glycine during 10 min. Cells were filtered and washed with PBS pH 7.2 (0.14 M NaCl, 2.7 mM KCl, 10 mM PO_4^{3-}). The harvested material was ground in liquid nitrogen and solubilized in NLB buffer (50 mM HEPES pH 7.5, 150 mM NaCl, 0.5 mM EDTA, 1% Triton X-100, 0.1% sodium deoxycholate, 0.1% SDS, 1 µg/ml pepstatin A, 1 µg/ml aprotinin, 1 µg/ml leupeptin, 1 mM PMSF) with an Ultra-Turrax T25 mixer. Chromatin was fragmented on ice with a probe sonicator to obtain ~200–800-bp fragments. After sonication, the suspension was centrifuged twice for 20 minutes at 16000 rpm, the protein concentration was determined by using the Bio-Rad protein assay kit, and the extract was used for chromatin precipitation.

For traditional ChIP, extract containing 200 mg protein was precleared with 100 µl salmon sperm DNA/protein A agarose beads (Millipore) for 1h. The precleared supernatant was incubated overnight with 12 µl anti-E2Fa-antibody (Takahashi et al. 2008) or no antibody (control). Afterwards, the supernatant was mixed with 100 µl salmon sperm DNA/protein A

agarose beads and incubated for 3 h on a rotating wheel. Finally, the beads were washed with 10 ml NLB buffer, 1 ml low salt buffer (20 mM TrisCl pH 8, 150 mM NaCl, 2 mM EDTA, 1% Triton X-100, 0.1% SDS), 1 ml high salt buffer (20 mM TrisCl pH 8, 500 mM NaCl, 2 mM EDTA, 1% Triton X-100, 0.1% SDS), and 1 ml LiCl buffer (10 mM TrisCl pH8, 0.25 M LiCl, 1 mM EDTA, 1% sodium deoxycholate, 1% NP-40).

For ChAP, wild-type and *E2Fa-HBH* overexpressing extract containing 200 mg protein was adjusted to 10 mM imidazole and incubated overnight on a rotating wheel with 100 µl Streptavidin Sepharose (GE Healthcare). Beads were washed with 10 ml NLB + 10 mM imidazol and three times with 1ml NLB containing 750 mM NaCl and increasing SDS concentrations (0.1%, 0.5% and 1%).

For TChAP, extract containing 200 mg protein was adjusted to 10 mM imidazole and incubated overnight on a rotating wheel with 300 µl Ni-NTA Superflow (Qiagen). The NiNTA resin was washed three times with 10 ml NLB + 10 mM imidazole and complexes were eluted with two times 1.5 ml NLB + 150 mM imidazole. Next, the eluates from three NiNTA purifications were pooled on 100 µl Streptavidin Sepharose. After 3-h incubation on a rotating wheel, the Streptavidin Sepharose beads were washed with 10 ml NLB + 10 mM imidazol and three times with 1ml NLB containing 750 mM NaCl and increasing SDS concentrations (0.1%, 0.5% and 1%).

Finally, the ChIP, ChAP and TChAP beads were washed with 15 ml TE buffer (10 mM TrisCl pH8, 500 mM NaCl, 1 mM EDTA) and the bound E2Fa-DNA complexes eluted and reverse crosslinked by incubation overnight at 65°C on a rotating wheel in 1 ml 10 mM TrisCl pH8, 1 mM EDTA, 0.5 M NaCl, 1% SDS + 0.5 µl RNase A (100 mg/ml). Next, the eluates were incubated with 100 µg Proteinase K for 2 h at 42°C on a rotating wheel and the DNA extracted by phenol/chloroform/IAA followed by purification using Qiaquick PCR purification kit and DNA quantification by the Quant-iT dsDNA High Sensitivity kit (Invitrogen).

Illumina sequencing

ChIP DNA was end-repaired and A-tailed as described in the Illumina ChIP-seq standard protocol. It was then ligated to customized adapters, which contain a 5-base barcode, before the final library was obtained by PCR amplification with the standard Illumina primers. The library was verified by cloning and sequencing of a few constructs, then quantified by Qubit (Invitrogen). The run was performed for 38 cycles on an Illumina Genome Analyzer IIx using 10 pM final concentration of DNA. The sample sequences are attributed to each sample according to their bar code (first 5 bases) with one mismatch allowed. For pooled samples with only two different bases in their bar code, the selected reads in the datasets are screened for shared read (attributed to both samples at first). They are then removed from both datasets.

Therefore, the datasets provided contain only sequences attributed without ambiguity to only one sample.

Data analysis

Read mapping was done with BWA (Li and Durbin 2009) using default settings. Reads that did not map uniquely to one of the five chromosomes were discarded from further analysis. Mapping was done against the TAIR9 version of the *A. thaliana* genome (<http://arabidopsis.org>).

Peak calling was performed using MACS (Zhang et al. 2008) with default setting, except the MFOLD parameter, which was lowered to 6, and the effective genome size, which was set to 90 Mb. The MACS confidence value is defined as $-10 \cdot \log_{10}(\text{p-value})$. Sub-sampling was done without replacement from the set of reads uniquely mapping to the five chromosomes. The target ratio between IP and control reads is 3:1. The ratio 3:1 was derived as the closest ratio to all of the complete sequencing sets. The datasets were downsampled to 600,000 reads total (150,000 control reads, 450,000 IP reads) to accommodate the smallest dataset.

Peak annotation was performed by overlapping the peak summits reported by MACS with the TAIR9 annotation. To define majority genes, the peak annotations from the different sub-samples were combined and only those genes, together with their maximum peak score, recovered in 50% or more of the samples were retained. E2F motif analysis was performed with dna-pattern using the motif consensus sequence TTTSSCGC (SEQ ID NO: 1) defined by Naouar et al. (2009).

Based on Naouar et al. (2009), 9287 genes contained an E2F motif (TTTSSCGC) in their 1-kb promoter region (Figure 4). To calculate the E2F motif enrichment in the E2Fa bound regions (Figure 3), the background motif frequency (4.34%) was estimated using reshuffled sequences with the same length and nucleotide composition as the bound regions defined by MACS.

Functional Gene Ontology enrichment analysis was performed using PLAZA 2.0 (Proost et al. 2009).

Quantitative ChIP/ChAP/TChAP-PCR

qPCR experiments were performed on total, ChIP, ChAP and TChAP isolated genomic DNA from wild-type and *E2Fa-HBH* overexpressing PSB-D cultures. qPCR experiments were done as technical triplicates on equal concentrations of input, ChIP, ChAP or TChAP DNA using the Roche LightCycler 480 system and the LC480 SYBR Green I Master kit (Roche Diagnostics). Primers used for PCR amplification were designed surrounding the position of maximum height in the corresponding peak (i.e., the peak center) identified in the sequencing

experiments, using Primer-BLAST (<http://www.ncbi.nlm.nih.gov/tools/primer-blast/>). Normalization of enrichment relative to total input DNA was calculated as $2^{(Ct(input)-Ct(purified))}$.

Example 1: Establishment of tandem chromatin affinity purification (TChAP) in *Arabidopsis* cell suspension cultures

Although ChIP has been applied to a wide range of model organisms, including *Arabidopsis*, it remains a challenging technique with the quality of specific antibodies greatly influencing genomic ChIP outcomes. Chromatin affinity purification (ChAP) methods, with different generic tags, are commonly used as an alternative, but also these approaches often suffer from low signal-to-noise ratios. To overcome this problem and to reduce nonspecific background we developed a two-step tandem chromatin affinity purification (TChAP) strategy. In addition, *Arabidopsis* cell suspension cultures, rather than plants, were utilized in order to have an unlimited supply of cell numbers.

Previous reports described a tandem affinity tag system, based on a biotinylated peptide flanked by two hexahistidine tags (HBH tag), compatible with TAP purification of formaldehyde crosslinked protein complexes in yeast (Guerrero et al. 2006; Tagwerker et al. 2006). To evaluate the suitability of this HBH tag for chromatin isolation in plants, we C-terminally tagged the well characterized *Arabidopsis* E2Fa TF and expressed the E2Fa-HBH fusion under control of the constitutive Cauliflower Mosaic virus (CaMV) 35S promoter in *Arabidopsis* cell suspension cultures. E2Fa-HBH expression in transgenic lines was determined by protein blot analysis with both an anti-His and an anti-E2Fa antibody (Takahashi et al. 2008)(Fig. 1A). Despite constitutive overexpression, E2Fa-HBH protein accumulation was close to the endogenous E2Fa protein level (Fig. 1A).

In vivo biotinylation of the HBH tag was initially assayed in total protein extracts using a Streptavidin-HRP conjugate detecting a biotinylated protein corresponding to E2Fa-HBH in transgenic extracts (Fig. 1A). In addition, biotinylation efficiency, as evaluated by incubating E2Fa-HBH protein extracts with Streptavidin Sepharose beads, was estimated at 85% when comparing the E2Fa-HBH amount in the bound versus the total fraction (Fig. 1B, see Methods).

While establishing the TChAP procedure, several aspects of the assay were optimized. Crosslinking stringency was evaluated to ensure good extraction and isolation of TF-DNA complexes; the suitable sonication condition was determined to obtain fragment lengths between 200-800 bp; and a purification strategy was set up to obtain a purification efficiency comparable with that of traditional TAP purifications, resulting in an E2Fa-HBH protein bait recovery of 2-5% (Fig. 2A).

Eventually TChAP was performed by combining the HBH purification method of Tardiff et al. (2007) with ChIP protocol de-crosslinking, de-proteinization and DNA purification (Kim et al.

2008). Briefly, E2Fa-HBH and its crosslinked proteins and DNA were first bound on NiNTA beads, specifically eluted with imidazole, and then bound to Streptavidin Sepharose under high stringency conditions. E2Fa-HBH-DNA complexes were subsequently eluted and reverse crosslinked and the DNA was purified.

Final evaluation of the TChAP procedure occurred by analyzing the TChAP DNA sample by quantitative PCR. The proximal promoters of the well known E2Fa target genes *ORC1b*, *CTF18* and *ETG1* (Vandepoele et al. 2005) were found to be highly enriched in comparison with the promoter regions of the control genes *CDKA;1* and *UBQ10* (Fig. 2B), illustrating the suitability of our TChAP approach for the isolation and identification of TF-DNA binding sites.

Example 2: Evaluation of alternative chromatin isolation methods

To further validate the TChAP strategy, we performed, in parallel, ChIP and single chromatin affinity purification (ChAP) reactions on the *E2Fa-HBH* expressing *Arabidopsis* cell suspension culture. Using a specific anti-E2Fa antibody (Takahashi et al. 2008), both endogenous E2Fa and E2Fa-HBH crosslinked DNA elements were precipitated. Biotin-tagging in combination with Streptavidin-based chromatin affinity purification (ChAP), which had been already reported to be a suitable ChIP alternative (de Boer et al. 2003; Kim et al. 2008), was also performed on E2Fa-HBH.

Analysis of the E2Fa-HBH protein purification by protein blotting demonstrates a higher recovery of bait protein using these one-step purification methods (Fig. 2A). As expected, the very high affinity of biotin for streptavidin (10^{-15} M K_d) is more efficient in isolating tagged E2Fa (40-50% estimated recovery of bait) than the traditional immuno-affinity approach using the E2Fa antibody (10-15%), whereas two-step affinity purifications in general result in low yields (2-5%). In concordance, quantification of the E2Fa bound co-isolated DNA measured the highest DNA yield in ChAP (120 ng), an intermediate amount in ChIP (30 ng) and the lowest concentration in TChAP (5 ng).

Quantitative PCR evaluation of the three methods shows that TChAP however gives the highest relative enrichment of selected E2Fa target gene promoters (Fig. 2B), suggesting a considerable better signal-to-noise ratio. In conclusion, these comparisons demonstrate that the TChAP procedure offers an excellent alternative to isolate TF bound chromatin, omitting the need for high quality antibodies and utilizing biotin-tagging, which enhances ChIP sensitivity, in combination with a second tandem purification to further improve enrichment.

EXAMPLE 3: Genome-wide comparison of ChIP, ChAP and TChAP

To further benchmark the alternative chromatin isolation methods, we analyzed their DNA samples by high-throughput sequencing. The different isolated DNA samples and their corresponding negative controls, namely DNA samples of ChAP and TChAP performed on

wild-type cultures and a no antibody ChIP on the *E2Fa-HBH* expressing culture, were processed by standard Illumina ChIP-seq barcoded library generation protocols and sequenced by multiplexing on an Illumina GAll. For the different samples, between 161,071 and 2,178,951 sequencing reads were uniquely mapped to the five chromosomes of the TAIR9 genome.

E2Fa-HBH enriched binding sites were identified with the MACS peak calling tool (Zhang et al. 2008). To eliminate possible bias in peak calling outcome caused by the variable amounts of read data of the different samples, we used a random sub-sampling strategy, fixing the number of reads in the datasets and repeating the peak calling analyses 100 times, to establish the stability of the peak calling algorithm in detecting the same enriched locations. Starting with the same number of reads (see Methods) 4161, 2221 and 8365 enriched regions were identified and linked to nearest neighbor/closest annotated genes for ChIP, ChAP and TChAP, respectively (Table 1). Among these, only majority peaks/genes, which are peaks/genes detected in at least half of all sub-samples, and the 200 best peaks/genes, as reported by MACS, were retained as high-confidence E2Fa bound genes in subsequent analyses. Overall, the TChAP method yielded more peaks and more majority genes compared to the other methods (Table 1). Additionally, the MACS peak score, which is a measure for the confidence level of the called peaks, is on average the highest for TChAP (Table 1, Table 3). By selecting only the 200 best peaks of each chromatin isolation procedure, in order to remove the different number of called peaks, the difference in score and confidence level becomes even more pronounced in favor for TChAP (Table 1). When all mapped reads are used for MACS peak calling analysis, a similar trend is obtained with TChAP delivering more peaks/genes with better peak scores.

Table 1. Number of peaks called over 100 sub-samplings of the complete datasets.

	Total # peaks (all samples)	Average # peaks (per sample)	# Majority genes	Median peak score majority genes (TOP200 peaks)
ChIP	4161	624	415	229 (410)
ChAP	2221	997	921	126 (394)
TChAP	8365	2975	2598	313 (2263)

The top 200 genes (TOP200) identified by ChIP, ChAP and TChAP of E2Fa-HBH display a very big overlap (Figure 5). Comparison with previous datasets obtained by transcriptome

and cis-regulatory element analysis (Vandepoele et al. 2005; Naouar et al. 2009) demonstrates high enrichments for E2Fa regulated genes and genes containing an E2F motif in the bound regions (E2F motif genes). Together, the high overlap and enrichments support the correctness of the datasets and prove the quality of the three different chromatin isolation methods used (Figure 5).

When examining the ChIP, ChAP and TChAP majority genes, 98% (410/415) ChIP and 78% (724/921) ChAP genes are also present in the TChAP majority dataset (Fig. 3A). However, 1779 genes (68%) of the much larger group of TChAP majority genes is unique. The average confidence level (MACS scores) of this specific group of genes is higher than those of respectively all and common identified ChIP or ChAP genes, suggesting that the unique TChAP genes are *bona fide* E2Fa bound genes that are missed by ChIP and ChAP because of the lower specificity of these methods. Comparison of the read coverage between the methods, as visualized by GenomeView (Abeel et al. 2012), at the *ORC1b* and *CTF18* promoter E2Fa binding sites, confirms this asset of TChAP-seq (Fig. 3B). As a consequence, both E2Fa regulated (164 genes, 3-fold enrichment) and E2F motif genes (439 genes, 6-fold enrichment) are found among the TChAP specific majority genes (Fig. 3A). Moreover, plotting the ChIP, ChAP and TChAP majority peaks/genes in function of the MACS confidence score and measuring the fraction of E2Fa regulated and E2F motif genes reveals that, despite the pattern that lower confidence levels correlate with lower overlap scores, all TChAP genes retain enrichment for these gene sets (Fig. 3C). As an additional validation, E2Fa binding specificity to several genes, uniquely identified with TChAP, was confirmed by quantitative PCR after TChAP (Figure 6).

In summary, above results show that, although many E2Fa target genes are identified by all three methods, much more E2Fa bound genes are found in the TChAP-seq dataset because TChAP is more specific, allowing the identification of less prevalent E2Fa binding sites.

EXAMPLE 4: Validation of TChAP-seq E2Fa target genes

Detailed analysis of the distribution in the *Arabidopsis* genome of the TChAP-seq identified E2Fa binding sites and their annotation to genes is summarized in Table 2. Peak locations were determined in relation to the nearest annotated gene (TAIR9). 63% of all peaks were located 5' (intergenic or UTR), 24% resided in coding exons or introns, and 13% overlapped with 3' UTR or 3' intergenic regions (Fig. 4A, Table 2). Considering 5' promoter peaks, the median distance from a peak to the transcription start site was 236 bp, which is consistent with the role of E2Fa in classical transcriptional regulation. Of all peaks, 94% was assigned to a protein-coding gene and 91% had annotated features within 1000 bp of the E2F bound region. Overall, 2598 E2Fa bound majority genes were annotated.

Table 2. Summary of E2Fa TChAP-seq peak location and gene annotation.

Distribution	# Majority genes	# E2Fa regulated genes	# E2F motif genes
intergenic 5'	1229	266	338
coding 5'	363	44	107
UTR 5'	412	86	91
intron	58	9	21
coding 3'	193	9	56
UTR 3'	76	2	21
intergenic 3'	266	17	96
Total	2598	433	730

5

Combining this annotation with the transcript profiling data of *E2Fa-DPa* overexpressing plants (Naouar et al. 2009) indicates that 433 (17%) E2Fa bound genes show obvious E2F regulation (5-fold enrichment compared to the expected number of genes) (Table 2, Fig. 4B). 730 of all E2Fa bound regions (28%) contain an E2F motif (6-fold enrichment, Table 2) and a high-
10 confidence set of 196 E2Fa TChAP target genes contain both genomic properties of E2Fa regulation and motif presence (Fig. 4B). Comparison of the TChAP dataset with expression profiling data of *RBR* mutant plants (Gutzat et al. 2011), another component of the E2F/DP/RBR pathway, also displays a big overlap with 371 of the 600 shared genes being new TChAP genes (Fig. 4C). In addition, qRT-PCR analysis of the expression level of several putative E2Fa
15 targets, not identified in any of these transcriptome studies, demonstrates transcriptional regulation in *E2Fa-DPa* overexpressing *Arabidopsis* plants (De Veylder et al. 2002)(Figure 7).

Further validation of the TChAP datasets came from GO enrichment analyses. In agreement with E2Fa function, cell cycle (Van Leene et al. 2010), cell cycle periodic expressed (Menges et al. 2003) and DNA repair (GO:0006281) genes are strongly represented in the
20 total E2Fa TChAP dataset. When subtracting known E2Fa targets (Naouar et al. 2009), cell cycle related retain less, but still significant, enriched among the new TChAP identified genes (Fig. 4D).

Taken together, these analyses demonstrate the quality of the E2Fa TChAP-seq dataset and emphasize its strength. Not only many known E2Fa regulated genes got identified as
25 direct targets, but many relevant new E2Fa bound genes point to possible additional functions of the E2Fa transcription factor.

REFERENCES

- Abeel T, Van Parys T, Saeys Y, Galagan J, Van de Peer Y. 2012. GenomeView: a next-generation genome browser. *Nucleic Acids Res* **40**(2): e12.
- 5 Bowler C, Benvenuto G, Laflamme P, Molino D, Probst AV, Tariq M, Paszkowski J. 2004. Chromatin techniques for plant cells. *Plant J* **39**(5): 776-789.
- de Boer E, Rodriguez P, Bonte E, Krijgsveld J, Katsantoni E, Heck A, Grosveld F, Strouboulis J. 2003. Efficient biotinylation and single-step purification of tagged transcription factors in mammalian cells and transgenic mice. *Proceedings of the National Academy of Sciences of the United States of America* **100**(13): 7480-7485.
- 10 de Jager SM, Scofield S, Huntley RP, Robinson AS, den Boer BGW, Murray JAH. 2009. Dissecting regulatory pathways of G1/S control in *Arabidopsis*: common and distinct targets of CYCD3;1, E2Fa and E2Fc. *Plant molecular biology* **71**(4-5): 345-365.
- De Veylder L, Beeckman T, Beemster GTS, de Almeida Engler J, Ormenese S, Maes S, Naudts M, Van Der Schueren E, Jacqmard A, Engler G et al. 2002. Control of proliferation, endoreduplication and differentiation by the *Arabidopsis* E2Fa-DPa transcription factor. *EMBO J* **21**(6): 1360-1368.
- 15 Gendrel A-V, Lippman Z, Martienssen R, Colot V. 2005. Profiling histone modification patterns in plants using genomic tiling microarrays. *Nat Methods* **2**(3): 213-218.
- 20 Guerrero C, Tagwerker C, Kaiser P, Huang L. 2006. An integrated mass spectrometry-based proteomic approach - Quantitative analysis of tandem affinity-purified *in vivo* cross-linked protein complexes (QTAX) to decipher the 26 S proteasome-interacting network. *Mol Cell Proteomics* **5**(2): 366-378.
- Gutzat R, Borghi L, Fütterer J, Bischof S, Laizet Y, Hennig L, Feil R, Lunn J, Gruissem W. 2011. RETINOBLASTOMA-RELATED PROTEIN controls the transition to autotrophic plant development. *Development* **138**(14): 2977-2986.
- 25 Kaufmann K, Muiño JM, Østerås M, Farinelli L, Krajewski P, Angenent GC. 2010. Chromatin immunoprecipitation (ChIP) of plant transcription factors followed by sequencing (ChIP-SEQ) or hybridization to whole genome arrays (ChIP-CHIP). *Nat Protoc* **5**(3): 457-472.
- 30 Kim J, Chu J, Shen X, Wang J, Orkin SH. 2008. An extended transcriptional network for pluripotency of embryonic stem cells. *Cell* **132**(6): 1049-1061.
- Kim TH, Ren B. 2006. Genome-wide analysis of protein-DNA interactions. *Annu Rev Genom Hum G* **7**: 81-102.
- Kuo M-H, Allis CD. 1999. *In vivo* cross-linking and immunoprecipitation for studying dynamic protein:DNA associations in chromatin environment. *Methods* **19**(3): 425-433.
- 35

- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**(14): 1754-1760.
- Magyar Z, Horváth B, Khan S, Mohammed B, Henriques R, De Veylder L, Bakó L, Scheres B, Bögre L. 2012. *Arabidopsis* E2FA stimulates proliferation and endocycle separately through RBR-bound and RBR-free complexes. *EMBO J* **31**(6): 1480-1493.
- Menges M, Hennig L, Gruissem W, Murray JAH. 2003. Genome-wide gene expression in an *Arabidopsis* cell suspension. *Plant molecular biology* **53**(4): 423-442.
- Naouar N, Vandepoele K, Lammens T, Casneuf T, Zeller G, Van Hummelen P, Weigel D, Rätsch G, Inzé D, Kuiper M et al. 2009. Quantitative RNA expression analysis with Affymetrix Tiling 1.0R arrays identifies new E2F target genes. *Plant J* **57**(1): 184-194.
- Park PJ. 2009. ChIP-seq: advantages and challenges of a maturing technology. *Nat Rev Genet* **10**(10): 669-680.
- Parkinson NJ, Maslau S, Ferneyhough B, Zhang G, Gregory L, Buck D, Ragoussis J, Ponting CP, Fischer MD 2012. Preparation of high-quality next-generation sequencing libraries from picogram quantities of target DNA. *Genome Res* **22**(1): 125-133.
- Proost S, Van Bel M, Sterck L, Billiau K, Van Parys T, Van de Peer Y, Vandepoele K. 2009. PLAZA: a comparative genomics resource to study gene and genome evolution in plants. *Plant Cell* **21**(12): 3718-3731.
- Saleh A, Alvarez-Venegas R, Avramova Z. 2008. An efficient chromatin immunoprecipitation (ChIP) protocol for studying histone modifications in *Arabidopsis* plants. *Nat Protoc* **3**(6): 1018-1025.
- Tagwerker C, Flick K, Cui M, Guerrero C, Dou Y, Auer B, Baldi P, Huang L, Kaiser P. 2006. A tandem affinity tag for two-step purification under fully denaturing conditions: application in ubiquitin profiling and protein complex identification combined with *in vivo* cross-linking. *Mol Cell Proteomics* **5**(4): 737-748.
- Takahashi N, Lammens T, Boudolf V, Maes S, Yoshizumi T, De Jaeger G, Witters E, Inzé D, De Veylder L. 2008. The DNA replication checkpoint aids survival of plants deficient in the novel replisome factor ETG1. *EMBO J* **27**(13): 1840-1851.
- Tardiff DF, Abruzzi KC, Rosbash M. 2007. Protein characterization of *Saccharomyces cerevisiae* RNA polymerase II after *in vivo* cross-linking. *Proceedings of the National Academy of Sciences of the United States of America* **104**(50): 19948-19953.
- Valvekens D, Montagu MV, Van Lijsebettens M. 1988. *Agrobacterium tumefaciens*-mediated transformation of *Arabidopsis thaliana* root explants by using kanamycin selection. *Proceedings of the National Academy of Sciences of the United States of America* **85**(15): 5536-5540.

Van Leene J, Hollunder J, Eeckhout D, Persiau G, Van De Slijke E, Stals H, Van Isterdael G, Verkest A, Neiryck S, Buffel Y et al. 2010. Targeted interactomics reveals a complex core cell cycle machinery in *Arabidopsis thaliana*. *Mol Syst Biol* **6**: 397.

Van Leene J, Stals H, Eeckhout D, Persiau G, Van De Slijke E, Van Isterdael G, De Clercq A, Bonnet E, Laukens K, Remmerie N et al. 2007. A tandem affinity purification-based technology platform to study the cell cycle interactome in *Arabidopsis thaliana*. *Mol Cell Proteomics* **6**(7): 1226-1238.

Van Leene J, Witters E, Inzé D, De Jaeger G. 2008. Boosting tandem affinity purification of plant protein complexes. *Trends Plant Sci* **13**(10): 517-520.

Vandepoele K, Vlieghe K, Florquin K, Hennig L, Beemster GTS, Grissem W, Van de Peer Y, Inzé D, De Veylder L. 2005. Genome-wide identification of potential plant E2F target genes. *Plant Physiol* **139**(1): 316-328.

Vlieghe K, Vuylsteke M, Florquin K, Rombauts S, Maes S, Ormenese S, Van Hummelen P, Van de Peer Y, Inzé D, De Veylder L. 2003. Microarray analysis of *E2Fa-DPa*-overexpressing plants uncovers a cross-talking genetic network between DNA replication and nitrogen assimilation. *J Cell Sci* **116**(20): 4249-4259.

Zhang Y, Liu T, Meyer CA, Eeckhout J, Johnson DS, Bernstein BE, Nussbaum C, Myers RM, Brown M, Li W et al. 2008. Model-based Analysis of ChIP-Seq (MACS). *Genome Biol* **9**(9): R137.

CLAIMS

1. A method for isolating target genes for transcription factors in plants, comprising (1) transforming a plant cell with an expression vector comprising a transcription factor fused to at least two tags (2) expressing said transcription factor in the plant cell (3) crosslinking the proteins with the DNA (4) extracting the proteins and chromatin from the cell (5) fragmenting the chromatin (6) purifying the chromatin-protein complex using affinity binding towards the first tag and (7) purifying the fraction comprising the first tag using affinity binding towards the second tag.
2. The method according to claim 1, wherein at least one of the tags is selected from the group consisting of a polyhistidine tag and a biotine tag.
3. The method according to claim 2, wherein one tag is a polyhistidine tag and one tag is a biotine tag.
4. The method according to claim 2 or 3, wherein said affinity binding is selected from the group consisting of a nickel chelating carrier and a streptavidine carrier.

Figure 1

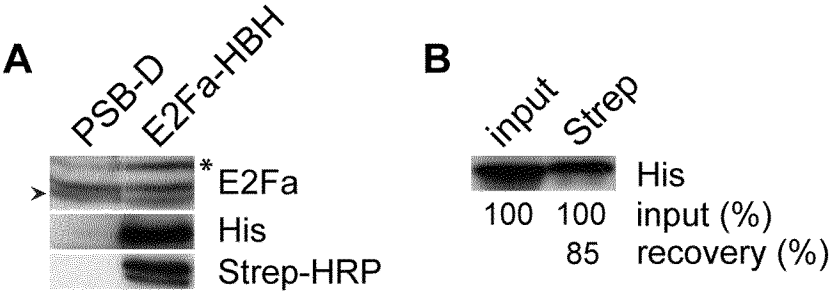


Figure 2

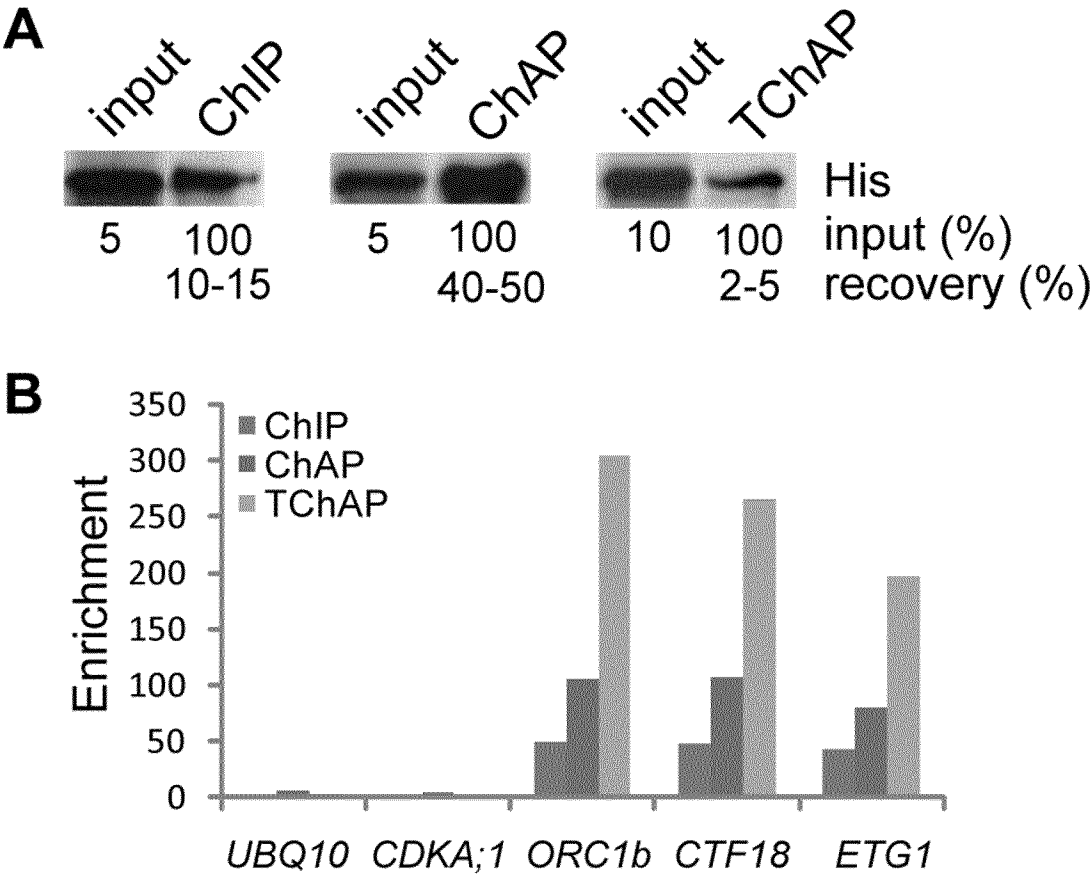


Figure 3

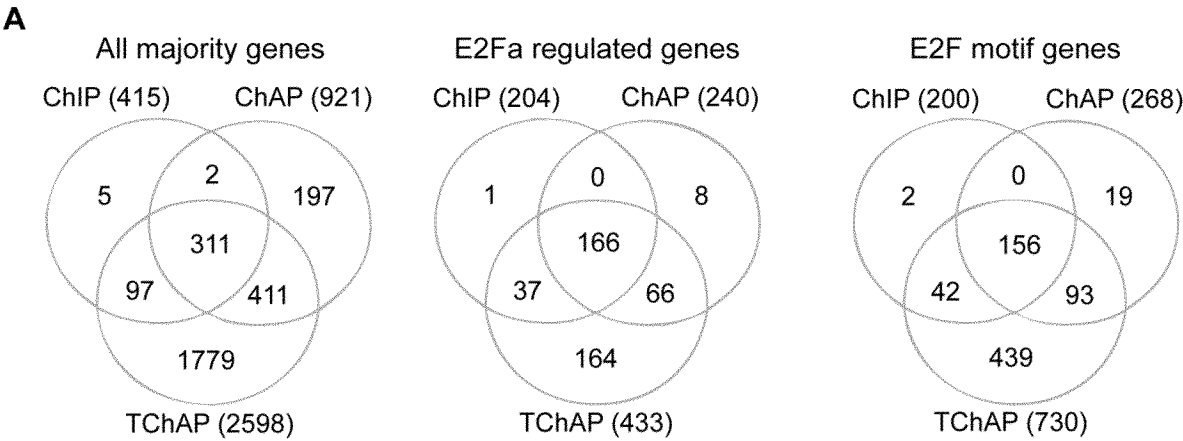


Figure 3 continued

B

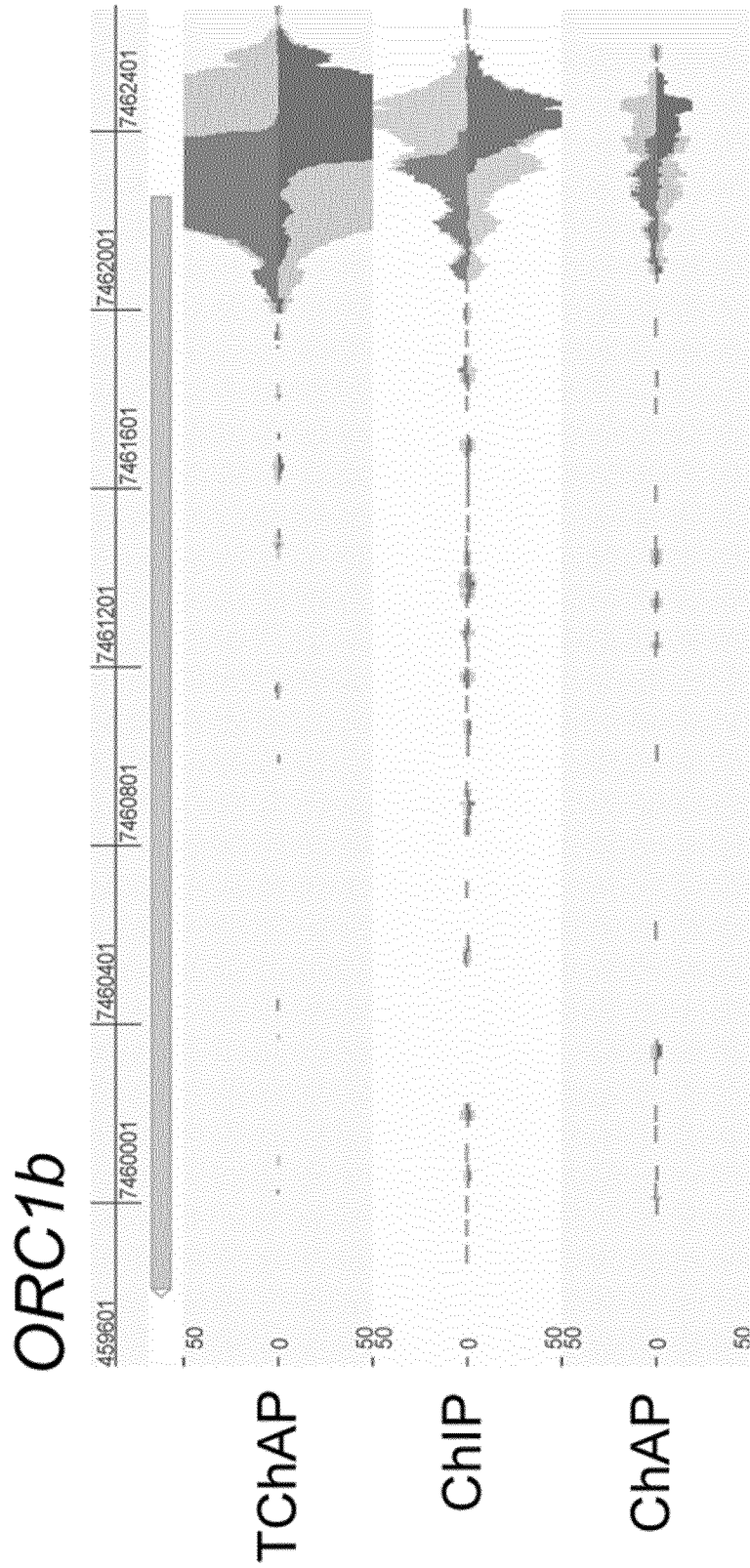


Figure 3 continued

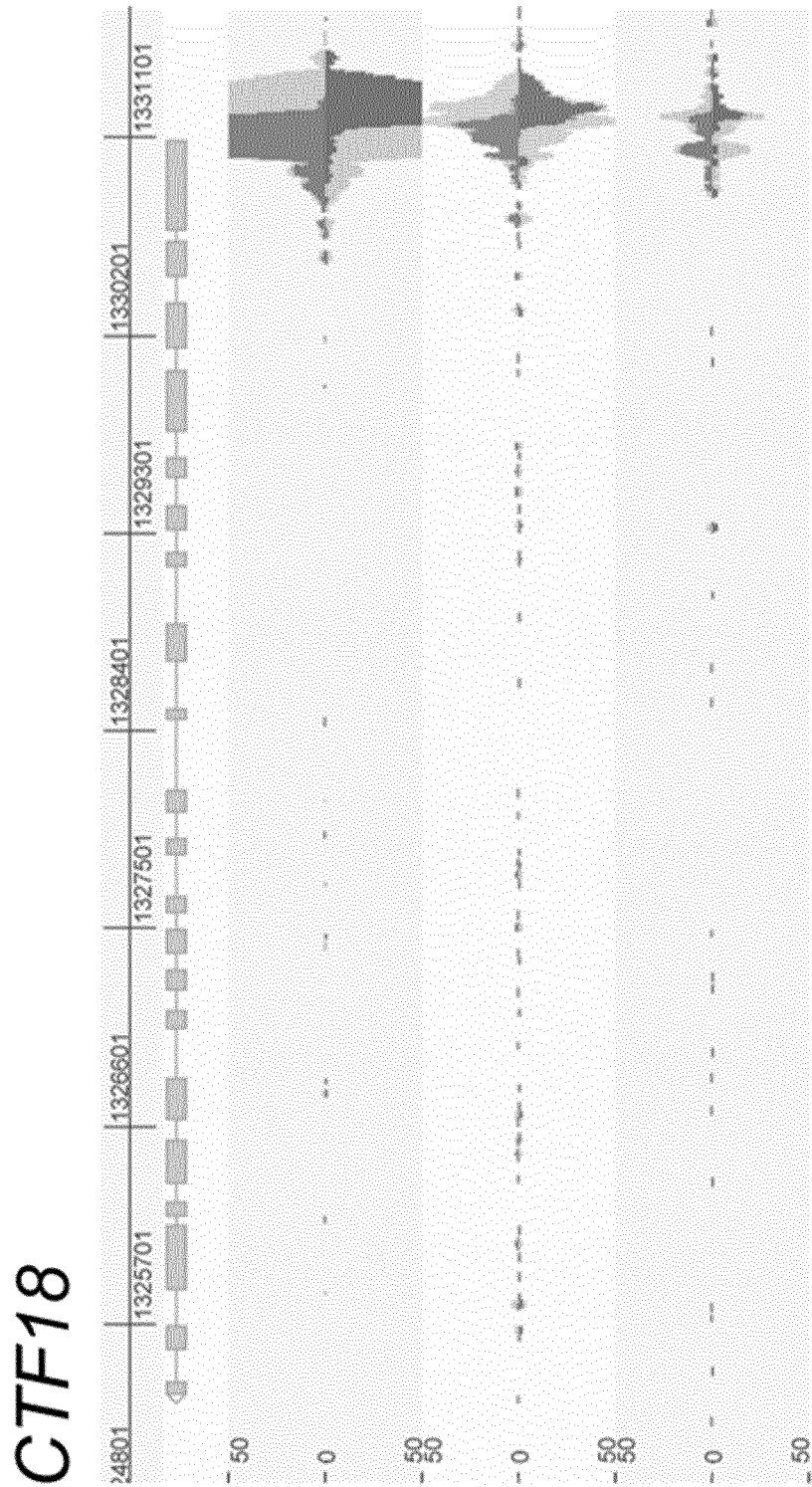


Figure 3 continued

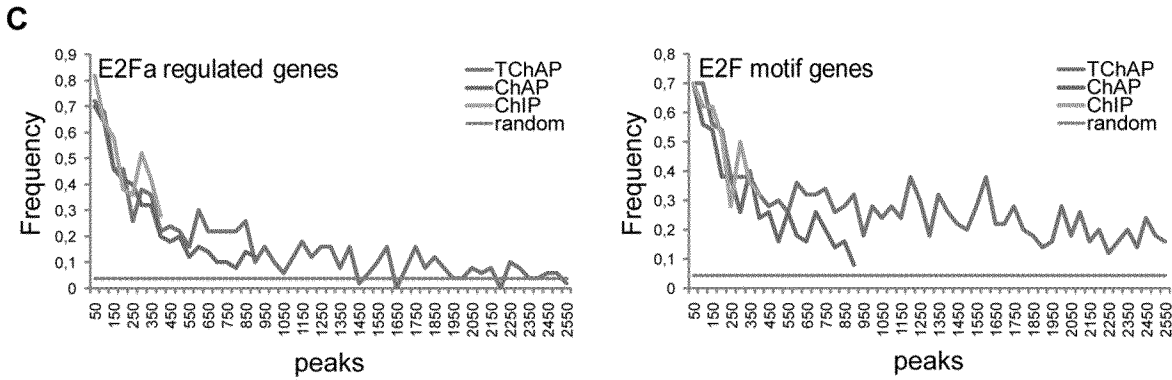


Figure 4

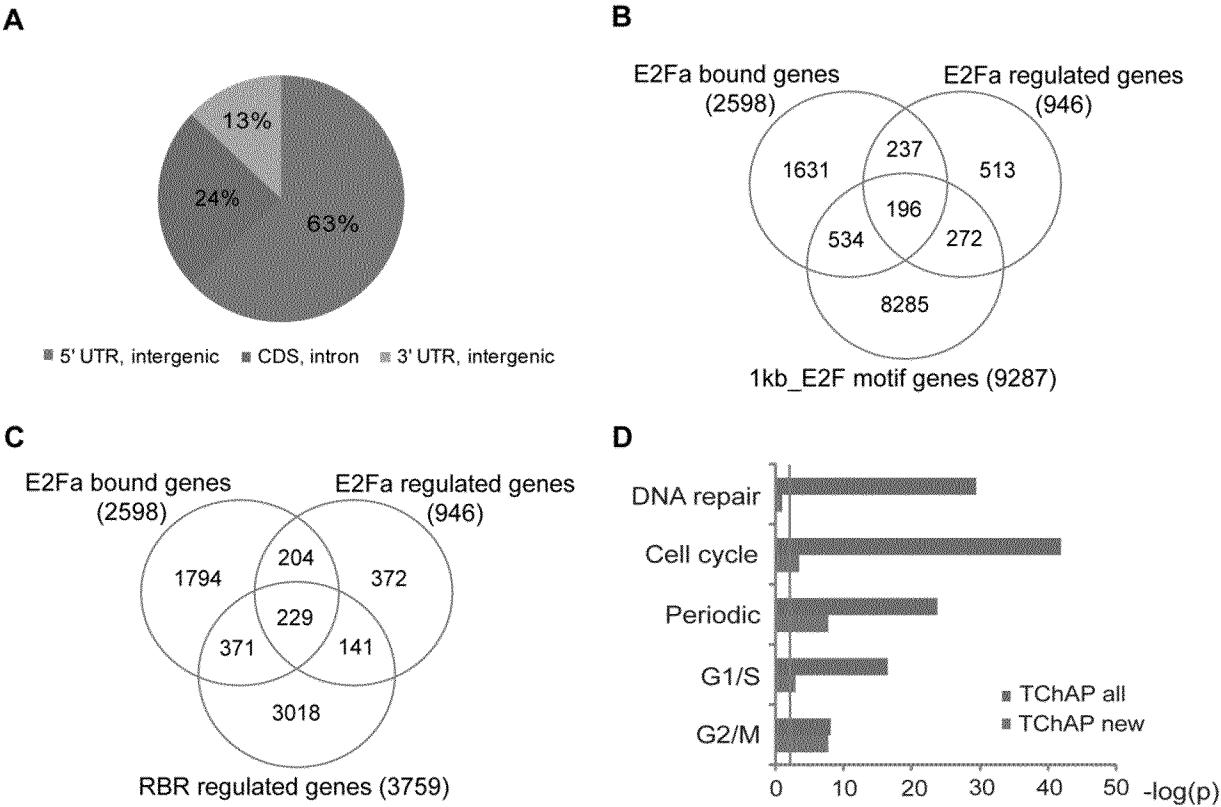


Figure 5

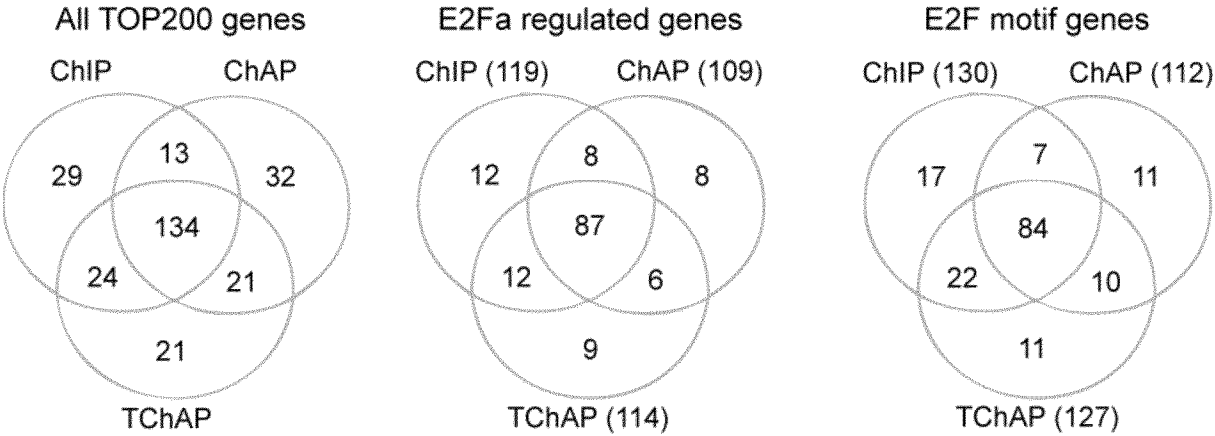


Figure 6

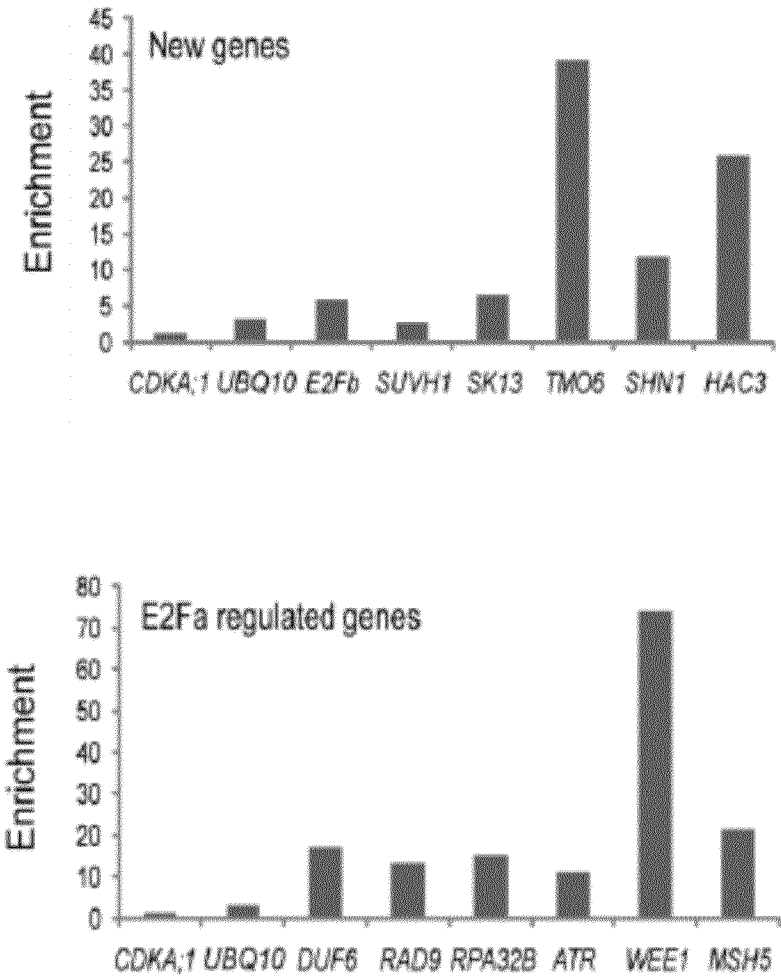


Figure 7

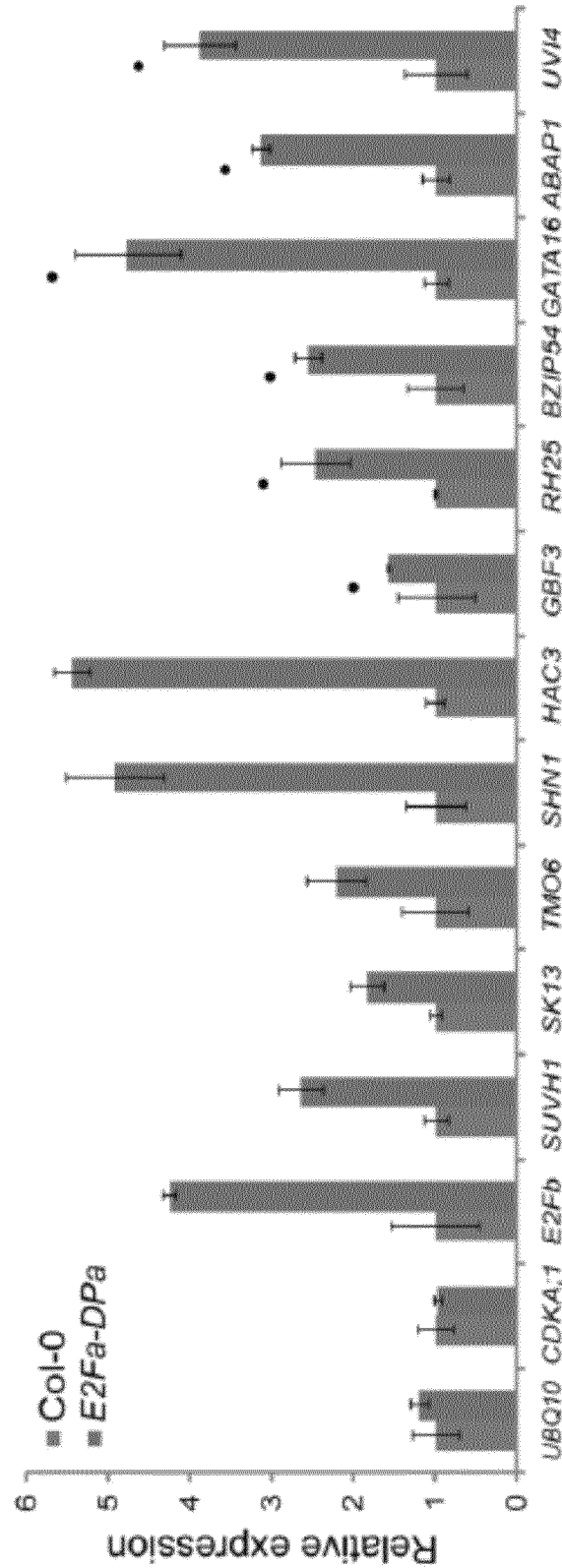


Figure 1

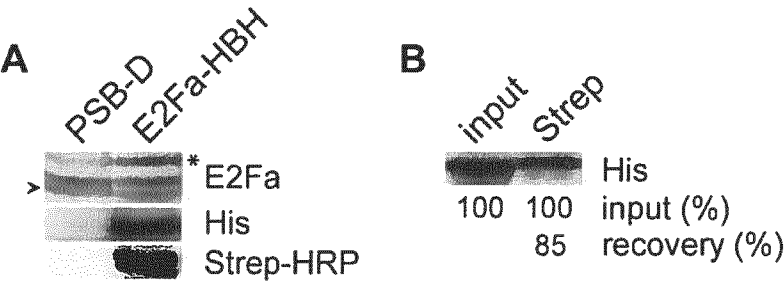


Figure 2

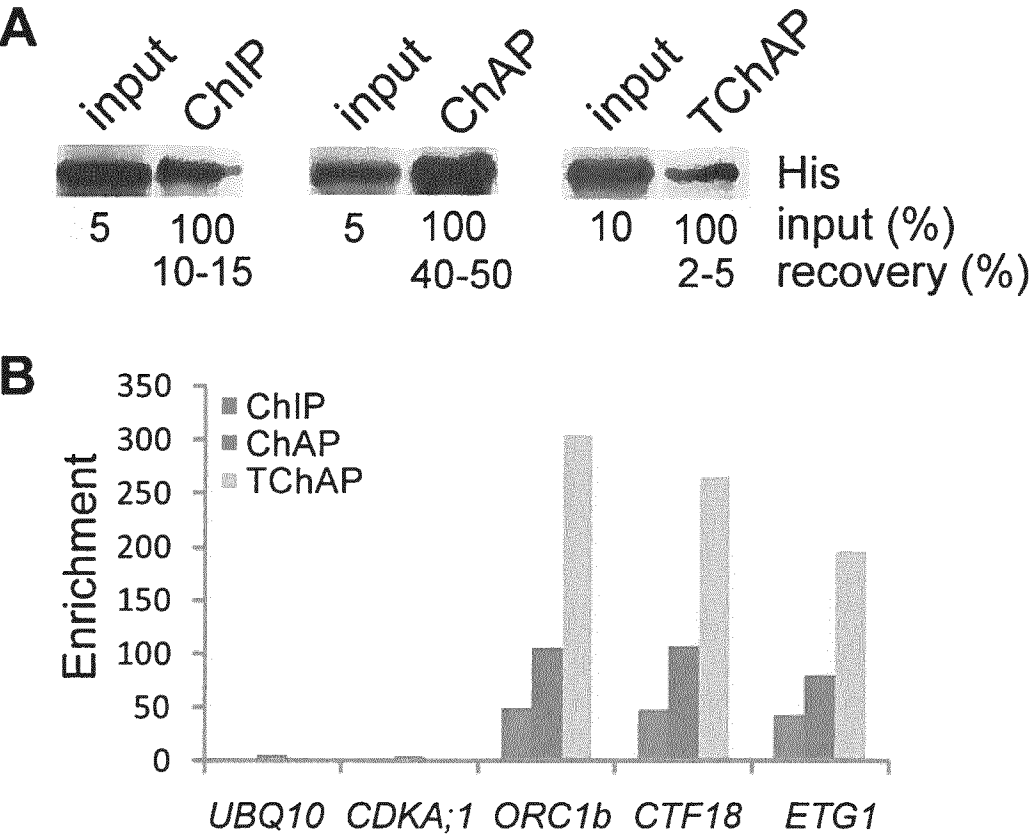


Figure 3

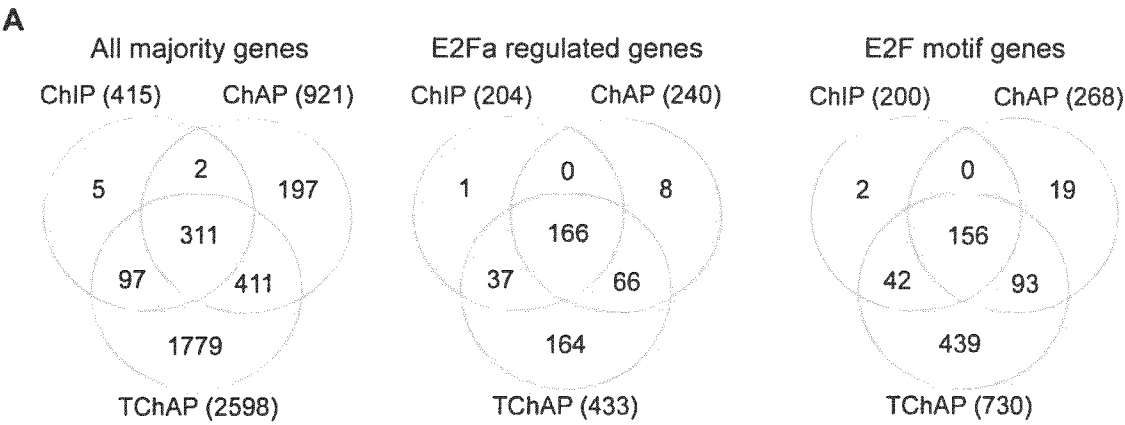


Figure 3 continued

B

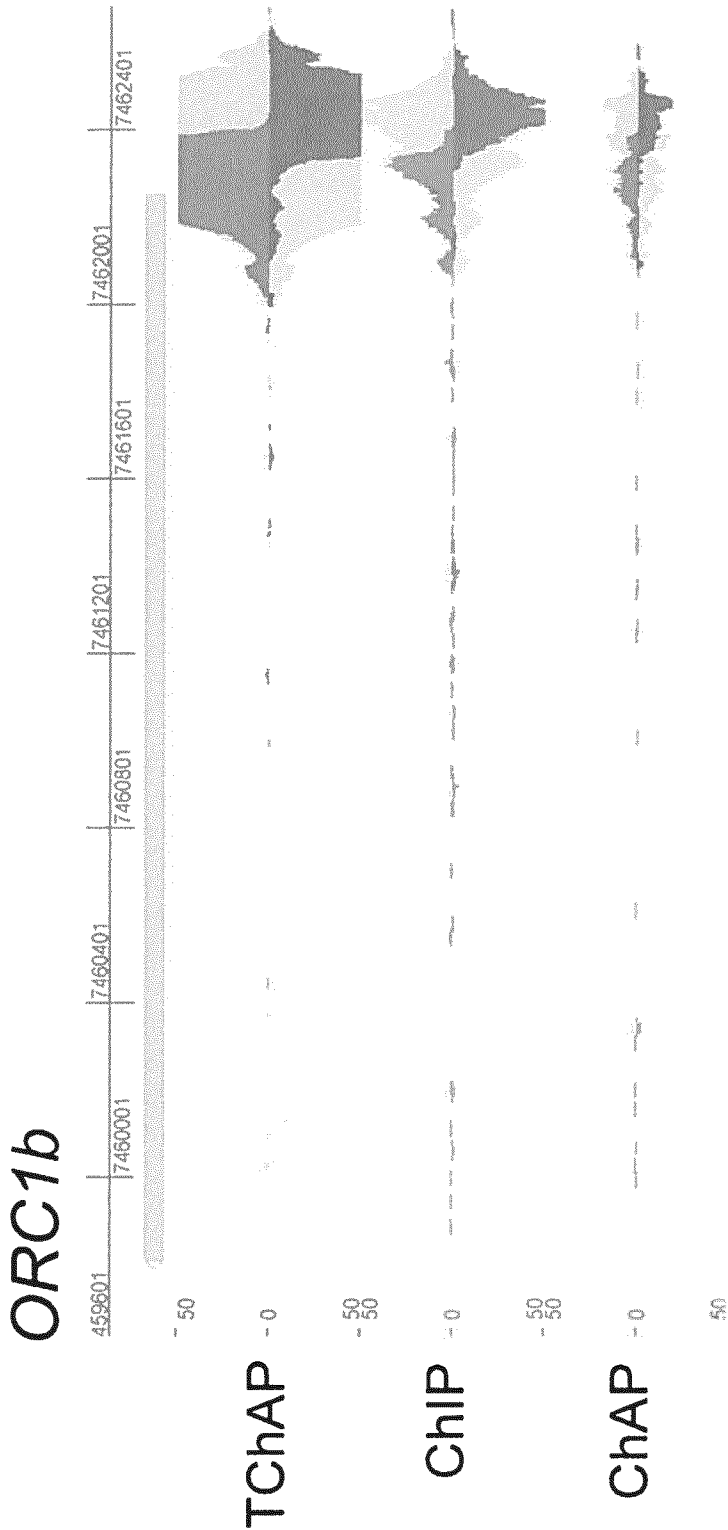


Figure 3 continued



Figure 3 continued

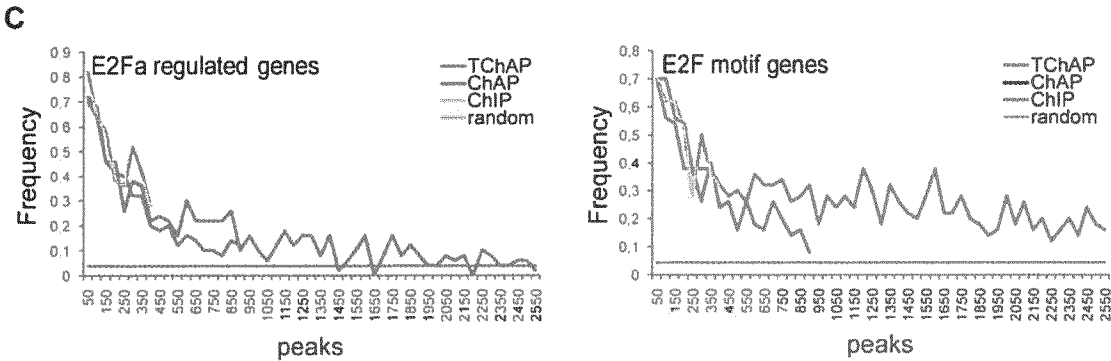


Figure 4

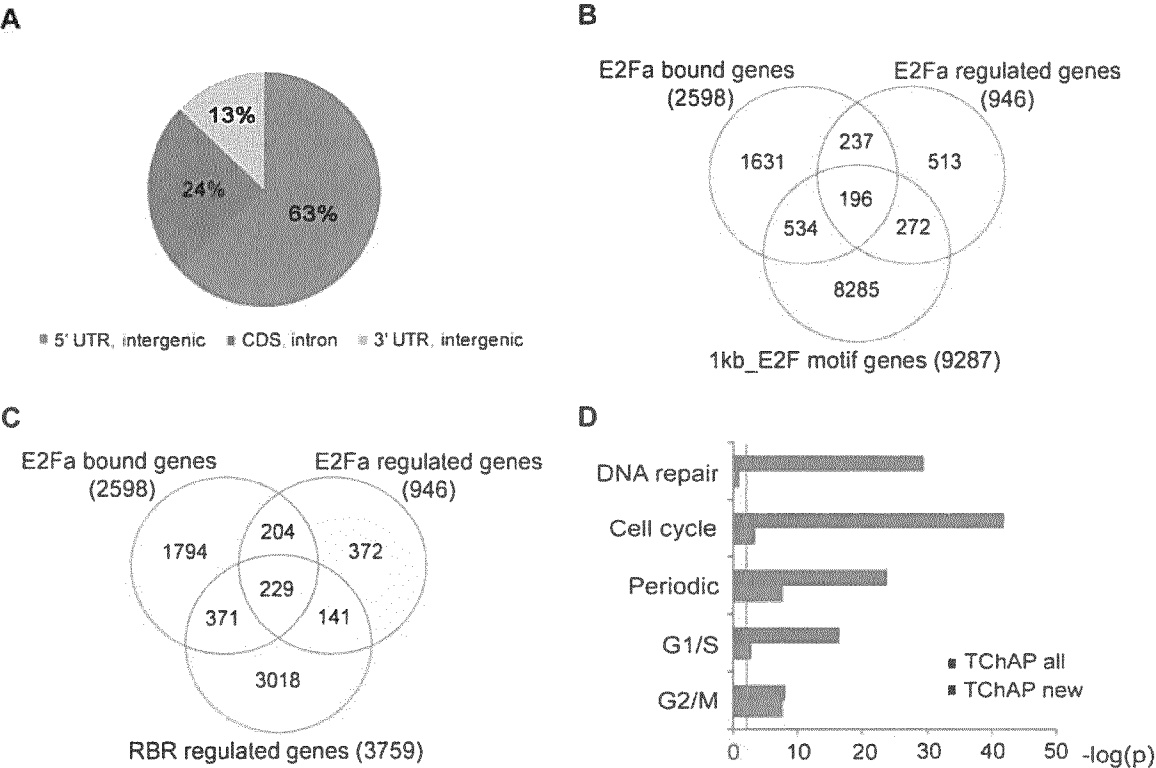


Figure 5

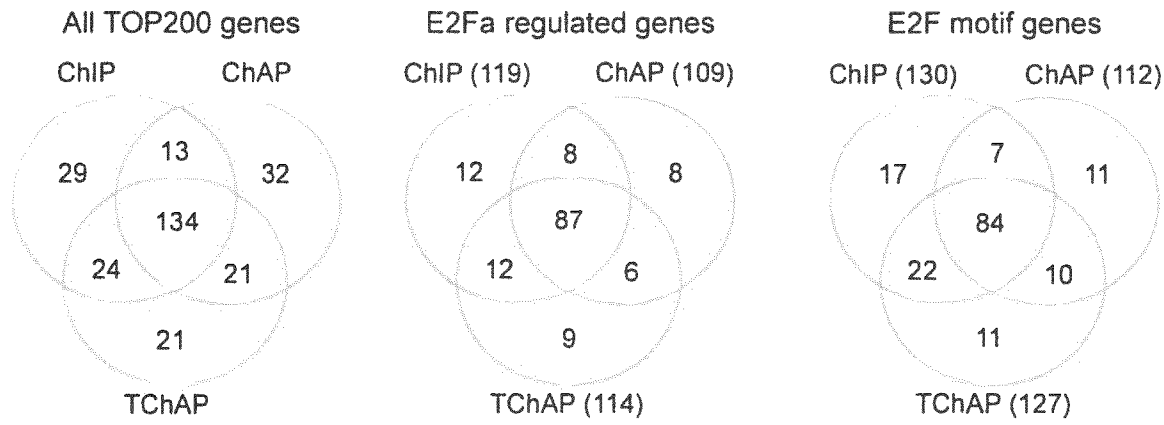


Figure 6

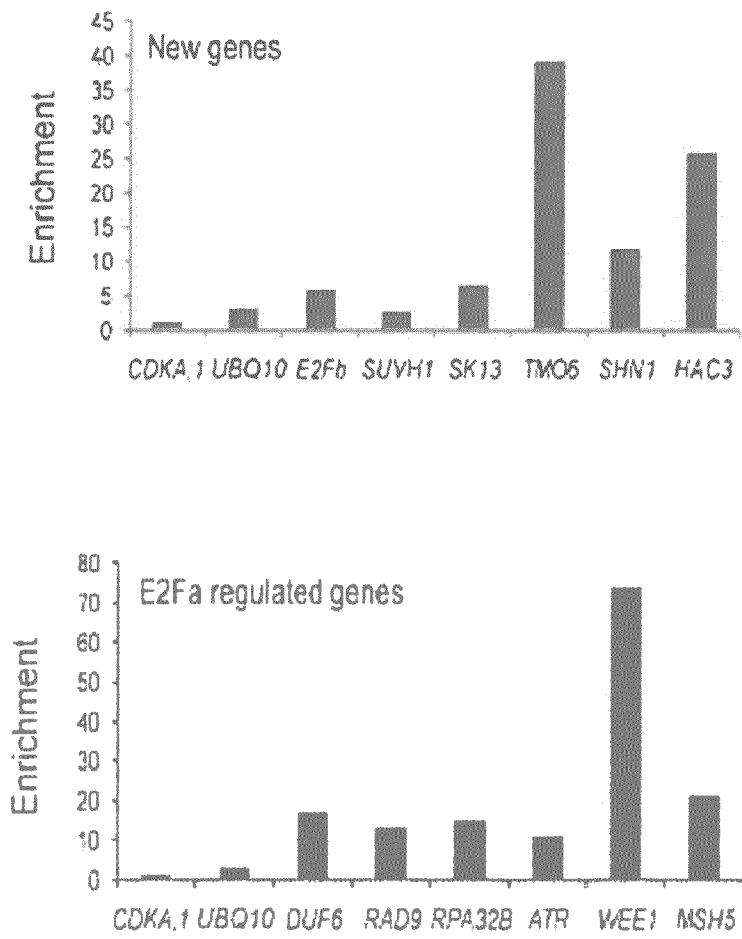
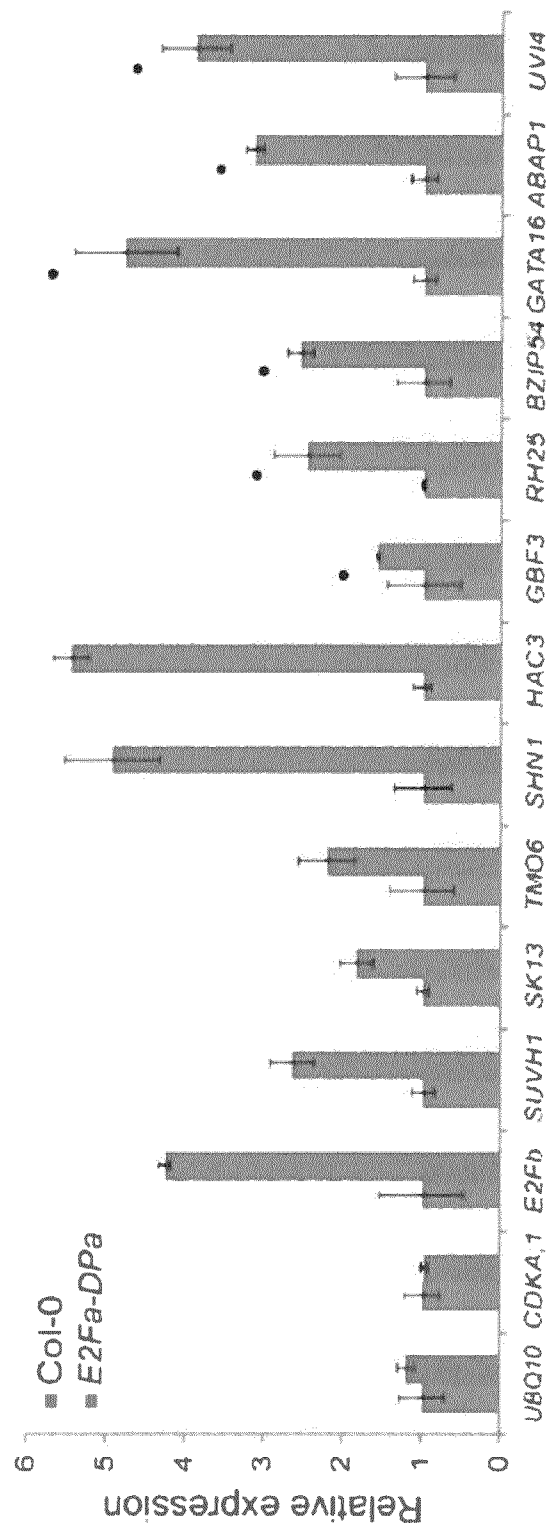


Figure 7



INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2013/065695

A. CLASSIFICATION OF SUBJECT MATTER

INV. C07K14/415 C12N15/62 C12N15/82 C12N15/10
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

C07K C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EPO-Internal, WPI Data, EMBASE, BIOSIS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	Verkest A; et al.: "Tandem chromatin affinity purification (TChAP) improves transcription factor target gene identification"	1
Y	In: Anonymous: "Comparative and Regulatory Genomics in Plants - Book of Abstracts", 12 April 2011 (2011-04-12), XP002712894, pages 1-32, page 21 ----- -/--	2-4



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

13 September 2013

Date of mailing of the international search report

02/10/2013

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Authorized officer

Wiame, Ilse

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2013/065695

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	C. TAGWERKER: "A Tandem Affinity Tag for Two-step Purification under Fully Denaturing Conditions: Application in Ubiquitin Profiling and Protein Complex Identification Combined with in vivoCross-Linking", MOLECULAR & CELLULAR PROTEOMICS, vol. 5, no. 4, 6 December 2005 (2005-12-06), pages 737-748, XP055079074, ISSN: 1535-9476, DOI: 10.1074/mcp.M500368-MCP200 cited in the application the whole document	2-4
A	Y. LI ET AL: "Absolute Quantitation of Isoforms of Post-translationally Modified Proteins in Transgenic Organism", MOLECULAR & CELLULAR PROTEOMICS, vol. 11, no. 8, 1 August 2012 (2012-08-01), pages 272-285, XP055079079, ISSN: 1535-9476, DOI: 10.1074/mcp.M111.016568 page 282, column 1, line 52 - column 2, line 26	2-4
A	R. HARADA ET AL: "Genome-wide location analysis and expression studies reveal a role for p110 CUX1 in the activation of DNA replication genes", NUCLEIC ACIDS RESEARCH, vol. 36, no. 1, 13 November 2007 (2007-11-13), pages 189-202, XP055078931, ISSN: 0305-1048, DOI: 10.1093/nar/gkm970 abstract; figure 1A page 191, column 2, paragraph 2	1
A	R. ZENTELLA ET AL: "Global Analysis of DELLA Direct Targets in Early Gibberellin Signaling in Arabidopsis", THE PLANT CELL ONLINE, vol. 19, no. 10, 26 October 2007 (2007-10-26), pages 3037-3057, XP055079028, ISSN: 1040-4651, DOI: 10.1105/tpc.107.054999 page 3045, column 1, last paragraph - page 3047, column 1, line 2; figure 4B	1,2