

RESEARCH ARTICLE

Open Access



# Profiling of chromatin accessibility identifies transcription factor binding sites across the genome of *Aspergillus* species

Lianggang Huang<sup>1</sup>, Xuejie Li<sup>1</sup>, Liangbo Dong<sup>1</sup>, Bin Wang<sup>1,2\*</sup> and Li Pan<sup>1,2\*</sup> 

## Abstract

**Background:** The identification of open chromatin regions and transcription factor binding sites (TFBs) is an important step in understanding the regulation of gene expression in diverse species. ATAC-seq is a technique used for such purpose by providing high-resolution measurements of chromatin accessibility revealed through integration of Tn5 transposase. However, the existence of cell walls in filamentous fungi and associated difficulty in purifying nuclei have precluded the routine application of this technique, leading to a lack of experimentally determined and computationally inferred data on the identity of genome-wide *cis*-regulatory elements (CREs) and TFBs. In this study, we constructed an ATAC-seq platform suitable for filamentous fungi and generated ATAC-seq libraries of *Aspergillus niger* and *Aspergillus oryzae* grown under a variety of conditions.

**Results:** We applied the ATAC-seq assay for filamentous fungi to delineate the syntenic orthologue and differentially changed chromatin accessibility regions among different *Aspergillus* species, during different culture conditions, and among specific TF-deleted strains. The syntenic orthologues of accessible regions were responsible for the conservative functions across *Aspergillus* species, while regions differentially changed between culture conditions and TFs mutants drove differential gene expression programs. Importantly, we suggest criteria to determine TFBs through the analysis of unbalanced cleavage of distinct TF-bound DNA strands by Tn5 transposase. Based on this criterion, we constructed data libraries of the *in vivo* genomic footprint of *A. niger* under distinct conditions, and generated a database of novel transcription factor binding motifs through comparison of footprints in TF-deleted strains. Furthermore, we validated the novel TFBs *in vivo* through an artificial synthetic minimal promoter system.

**Conclusions:** We characterized the chromatin accessibility regions of filamentous fungi species, and identified a complete TFBs map by ATAC-seq, which provides valuable data for future analyses of transcriptional regulation in filamentous fungi.

**Keywords:** Filamentous fungi ATAC-seq, Chromatin accessibility, Footprints, Transcription factor binding sites, *Aspergillus* species

\* Correspondence: [btbinwang@scut.edu.cn](mailto:btbinwang@scut.edu.cn); [btlipan@scut.edu.cn](mailto:btlipan@scut.edu.cn)

<sup>1</sup>School of Biology and Biological Engineering, South China University of Technology, Guangzhou Higher Education Mega Center, Guangzhou 510006, China

Full list of author information is available at the end of the article



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

## Background

There have been 339 species identified in the *Aspergilli* genus [1], and they are of broad interest because of their industrial applications, importance as pathogens for animals and crops, and relevance for basic research. *Aspergillus* section Nigri is the fungal genus with the most sequenced genomes, as genomes from 27 species are already publicly available [2], and genomic analyses have led to a better understanding of fungal biology and improvements in the industrial use of these organisms [1, 3, 4]. Gene regulation is of major way by which fungi physiologically act in environmental circumstances and respond to changing conditions [5, 6]. However, in comparison with *Saccharomyces cerevisiae*, the knowledge of genome-wide regulatory mechanisms in *Aspergilli* is lagging. Transcription factors (TFs) are key regulators of biological processes that function by binding to transcriptional regulatory regions to control the expression of target genes. The *cis*-regulatory elements (CREs) related to TF binding play pivotal roles in the regulation of gene expression, and each TF recognizes a collection of similar DNA sequences, known as binding site motifs [7, 8]. Therefore, the ability to identify these CREs and TF motifs throughout *Aspergillus* genomes is an important step in understanding the regulatory functions of TF binding and gene expression. Analysis of genome sequences has revealed an average of 600 putative transcription factors in each *Aspergillus* fungus genome [3, 9]. Nevertheless, approximately 5% of the TFs in the *Aspergillus* genus have only been identified and not been studied further [3, 9]. Most genome-wide TFs in the *Aspergillus* genome have not been experimentally determined or computationally inferred, and the sites remain unknown.

The gold-standard method for identifying *in vivo* CREs for TFs of interest is chromatin immunoprecipitation DNA sequencing (ChIP-seq) [10]. However, the scarcity of sequence-specific antibodies for most *Aspergillus* TFs or tagged target proteins for the more than 600 TFs in *Aspergillus* has prevented the widespread application of this method in *Aspergillus* [11]. So far, only a few ChIP-seq experiments have used epitope-tagged or customized TF antibodies to study *Aspergillus* TFs, including CrzA [12], SrbA [13], MetR [14], and AtrR [15] of *Aspergillus fumigatus*; WetA of *Aspergillus nidulans* [16]; CRE-1 of *Neurospora crassa* [17]; Tri6 of *Fusarium graminearum* [18]; MoCRZ1 of *Magnaporthe oryzae* [19] and some chromatin modification such as H3K4 trimethylation in *A. nidulans* [20, 21]. These findings reveal that the available filamentous fungi ChIP-seq data are not abundant and performing ChIP-seq experiments in filamentous fungi is full of challenges. Additional high-throughput sequencing methods, i.e., DNase-seq and formaldehyde-assisted isolation of regulatory elements with sequencing

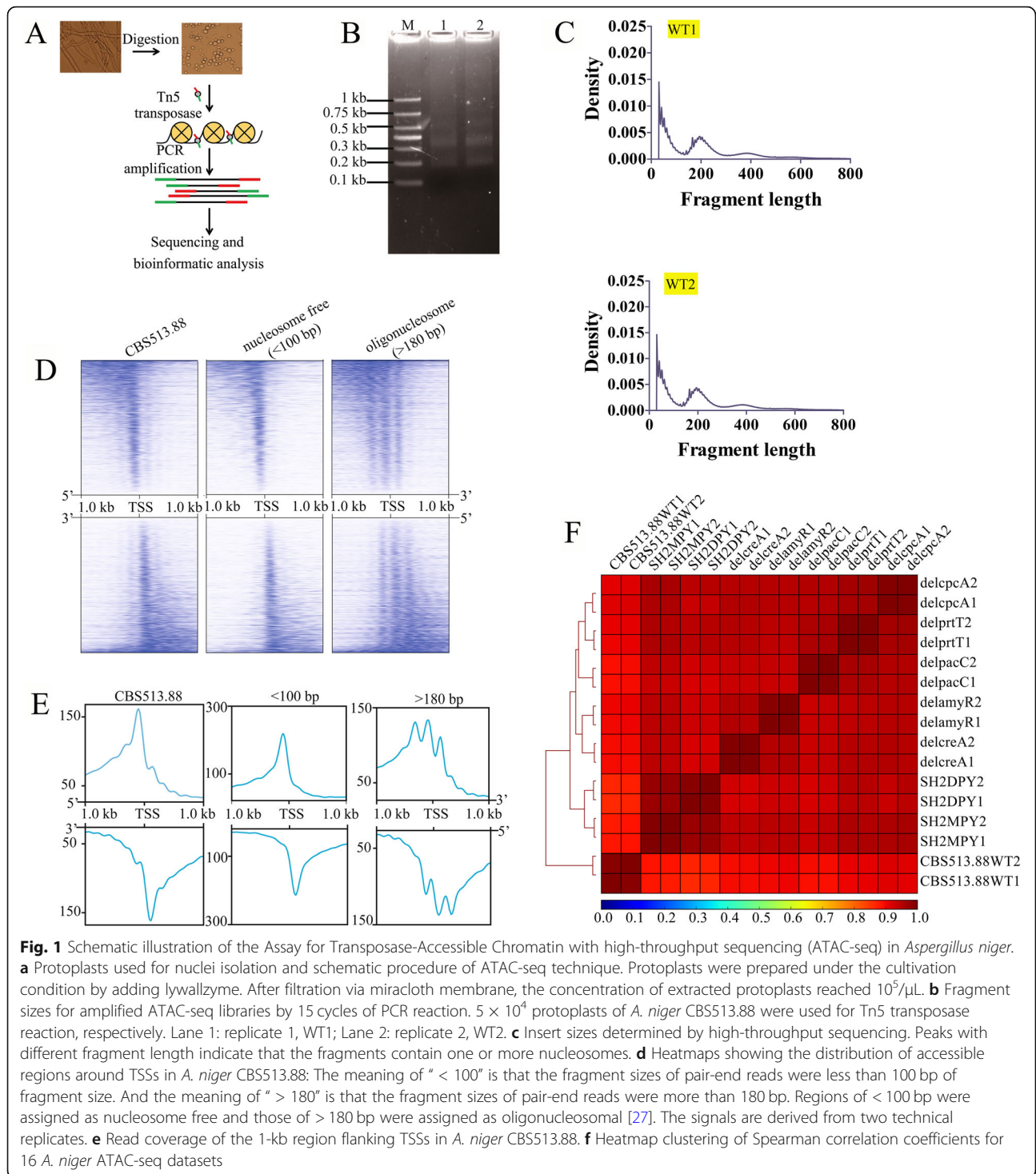
(FAIRE-seq) combine enzymatic digestion of isolated chromatin with sequencing, were developed to identify CREs on a genome-wide scale [22, 23]. The shortcomings of these methods are that they require millions of cells as starting material, complex and time-consuming sample preparations, and many potentially loss-prone steps, such as adaptor ligation, gel purification, and cross-link reversal [24]. For these reasons, the DNase-seq analysis that we previously carried out in *Aspergillus oryzae* could not be effectively replicated [23]. Therefore, the development of feasible and scalable methods is required to facilitate the identification of regulatory elements.

Assay for transposase accessible chromatin sequencing (ATAC-seq) is a recently developed technique used to identify accessible regions and DNA footprints [25, 26]. Tn5 transposase integrates sequencing adapters directly into DNA, eliminating the need for multiple reactions and purification steps typically required for the construction of a sequencing library. As a result, significantly lower amounts of starting nuclei are required for the investigation of CREs. ATAC-seq is now routinely being applied to systematically identify *cis*-regulatory regions and DNA footprints in humans [27], mice [28], zebrafish [29], *Drosophila* [30], *Caenorhabditis elegans* [31], *Arabidopsis thaliana* [32], *S. cerevisiae* [33], and so on. However, in *Aspergillus* species, the existence of cell walls and the purification of nuclei preparation prevented the routine application of this technique. To overcome this obstacle, we performed an ATAC-seq assay for *Aspergilli* by protoplast preparation under cultivation conditions. We combined ATAC-seq and RNA-seq of different *Aspergillus* species and TF mutants to profile chromatin accessibility and identify gene regulatory elements *in vivo* on a genome-wide scale. Furthermore, we characterized the identified CREs *in vivo* by the synthetic minimal promoter driving expression of the reporter gene. Collectively, this study is an important step towards the meticulous analysis of industrial microbial transcriptional regulatory networks and provides valuable resource for future research aimed at characterizing or using gene regulatory elements.

## Results

### The ATAC-seq assay in the filamentous fungus

To identify chromatin accessibility and regulatory elements in filamentous fungi, we developed an ATAC-seq protocol for *Aspergillus* species by adding a step that released native nuclei by detergent lysis of protoplasts prior to the transposition step (Fig. 1a). A total of  $5 \times 10^4$  protoplasts were subjected to the extraction of native nuclei in lysis buffer containing 0.05% IGEPAL CA-630 (Sigma, I8896). After Tn5 transposition, PCR and electrophoresis were performed to detect the discrete



nucleic acid bands (Fig. 1b). The insert size distribution of sequenced fragments from *A. niger* chromatin had a clear periodicity of approximately 200 bp (Fig. 1c, Additional file 1: Figure S1), suggesting many fragments are protected by integer multiples of nucleosomes.

We analyzed the pattern of mapped ATAC-seq reads around the transcription start site (TSS) in *A. niger*

CBS513.88. Heatmaps showed the intensity of mapped ATAC-seq reads flanking 1-kb upstream and downstream of annotated TSSs in the AspGD database (Fig. 1d). Furthermore, separate heatmaps showed that fragments less than 100 bp clustered immediately upstream of TSSs throughout the *A. niger* genome. Fragments between 180 and 247 bp, corresponding to nucleosome

footprints, were depleted from TSSs throughout the genome. Overall, the identified accessible regions were mostly enriched around the TSSs (Fig. 1d), which is consistent with the fact that these regions usually contain active regulatory *cis*-elements. The signal intensity of accessible regions showed a clear depletion and periodic peaks as the distance from the TSS (Fig. 1e), which is consistent with the presence of a positioned nucleosome + 1 that serves as barrier, dictating the positioning of neighboring nucleosomes (especially nucleosome - 1). This phenomenon attenuates at further distances from the barrier [34]. Furthermore, using ATAC-seq for *A. niger*, we generated paired-end ATAC-seq libraries for 16 samples, and the libraries ranged from 35.9 to 104.9 million reads (Additional file 2: Table S1). The percentages of reads in two replicate samples of CBS513.88\_WT1 and CBS513.88\_WT2 that aligned to the *A. niger* CBS513.88 genome was 84.23% and 86.31%, respectively (Additional file 2: Table S1). The insert size distributions of sequenced fragments from 16 ATAC-seq libraries of *A. niger* had a similar periodicity (Additional file 1: Figure S1). Spearman correlation analysis was used to inspect the reproducibility of biological and technical replicates. Heatmap clustering of Spearman correlation coefficients for 16 *A. niger* ATAC-seq datasets revealed a strong correlation between replicates of the same strain or the mutant and a lower correlation between different strains (Fig. 1f). Taken together, these data suggested that the ATAC-seq protocol of filamentous fungus *A. niger* could provide data on accessible regions of chromatin in a genome-wide search for regulatory elements.

#### Identification of genome-wide chromatin accessibility and comparison of open chromatin profiles among *Aspergillus* species and strains

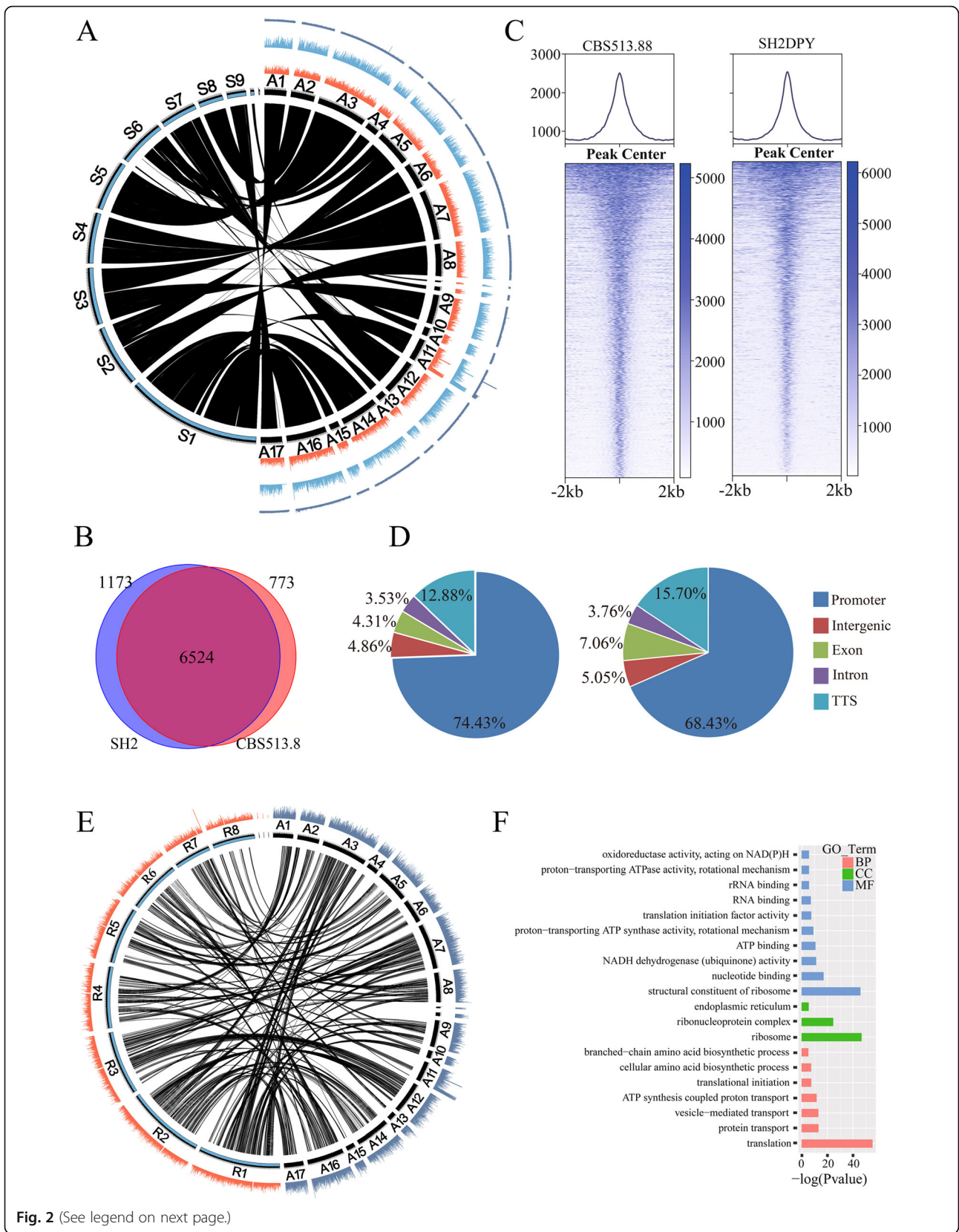
Visualization of the *A. niger* CBS513.88 and SH2 ATAC-seq data sets on a genome scale revealed that they were similar to each other, and the Tn5 integration regions of naked genomic DNA were uniformly distributed along the entire chromosome (Fig. 2a). The syntenic analysis showed that most of the genomic sequences were homologous between *A. niger* strains CBS513.88 and SH2, except for a missing DNA fragment in *A. niger* CBS513.88 contig VIII\_An18 and a fragment in *A. niger* SH2 contig S6 (Fig. 2a). A total of 7297 and 7697 accessible chromatin regions (peaks) were identified in *A. niger* CBS513.88 and SH2, respectively (Fig. 2b and Additional file 3: Table S2). Among these accessible regions, 6524 regions overlapped, indicating that these regions contained common *cis*-elements for both strains. These common regulatory elements might play a basal transcriptional regulatory role in *A. niger* species. Additionally, 773 and 1173 accessible regions are specific for *A.*

*niger* CBS513.88 and SH2, respectively, and these accessible regions contained differential *cis*-elements for each strain. A heatmap of the peak signal for the 2-kb flanking region of the peak center (Fig. 2c) showed that most of the accessible regions formed sharp peaks in the center and ranged from narrow to wide, which provided valuable data depending on *cis*-element motifs in these accessible regions. On the genome scale, accessible regions are highly enriched in promoter regions (Fig. 2d), which was consistent with the distribution of accessible regions in Fig. 1d. In addition to promoter regions, the transcription terminator site (TTS) regions contain the maximal accessible regions (12.88% for CBS513.88 and 15.70% for SH2, Fig. 2d). These accessible regions in TTSs indicated that gene downstream regions also contain regulatory elements for gene transcription.

As the GRAS strains, *A. niger* and *A. oryzae* share the common characteristics for the production of heterologous protein, protease, and natural products. To understand the syntenic orthologues of accessible regions across the *Aspergillus* species, we compared the *A. niger* CBS513.88 and *A. oryzae* niaD300 ATAC-seq peaks on a genome-wide scale. Our results showed that these two *Aspergillus* species had strong peak synteny, and there were 1564 syntenic orthologues of ATAC-seq peaks identified between *A. niger* CBS513.88 and *A. oryzae* niaD300 (Fig. 2e, Additional file 4: Table S3). Then, we located the target genes of these syntenic peaks and analyzed their GO enrichment function (Fig. 2f). Ribosomes related to GO categories were significantly enriched in these target genes, especially for the following three categories: “structural constituent of ribosome (MF),” “ribosome (CC),” and “translation (BP).” These highly enriched target genes indicated that *Aspergillus* species strains possessed the highly homologous regulatory mechanisms in terms of protein synthesis. Therefore, ATAC-seq data can be used for comparative genome analysis to explain the conservation of the evolutionary process of organisms at the genome level.

#### Identification of significantly changed chromatin accessibility driven by transcription factors and chromosome regulatory factor in *Aspergillus* species

A clustering map of ATAC-seq peaks for 16 *A. niger* samples indicates that there were common ATAC-seq patterns among all *A. niger* strains and specifically showed that SH2 and its derivatives did not have ATAC-seq signals corresponding to the 200-kb missing genomic DNA fragments (Fig. 3a) [4]. Furthermore, we identified the ATAC-seq peaks that significantly changed accessibility (FDR < 0.05) among *A. niger* SH2 cultured under different conditions and its TF-deleted strains (Fig. 3b and Additional file 5: Table S4). A total of 4009 of the differential ATAC-seq peaks (DAPs) in the



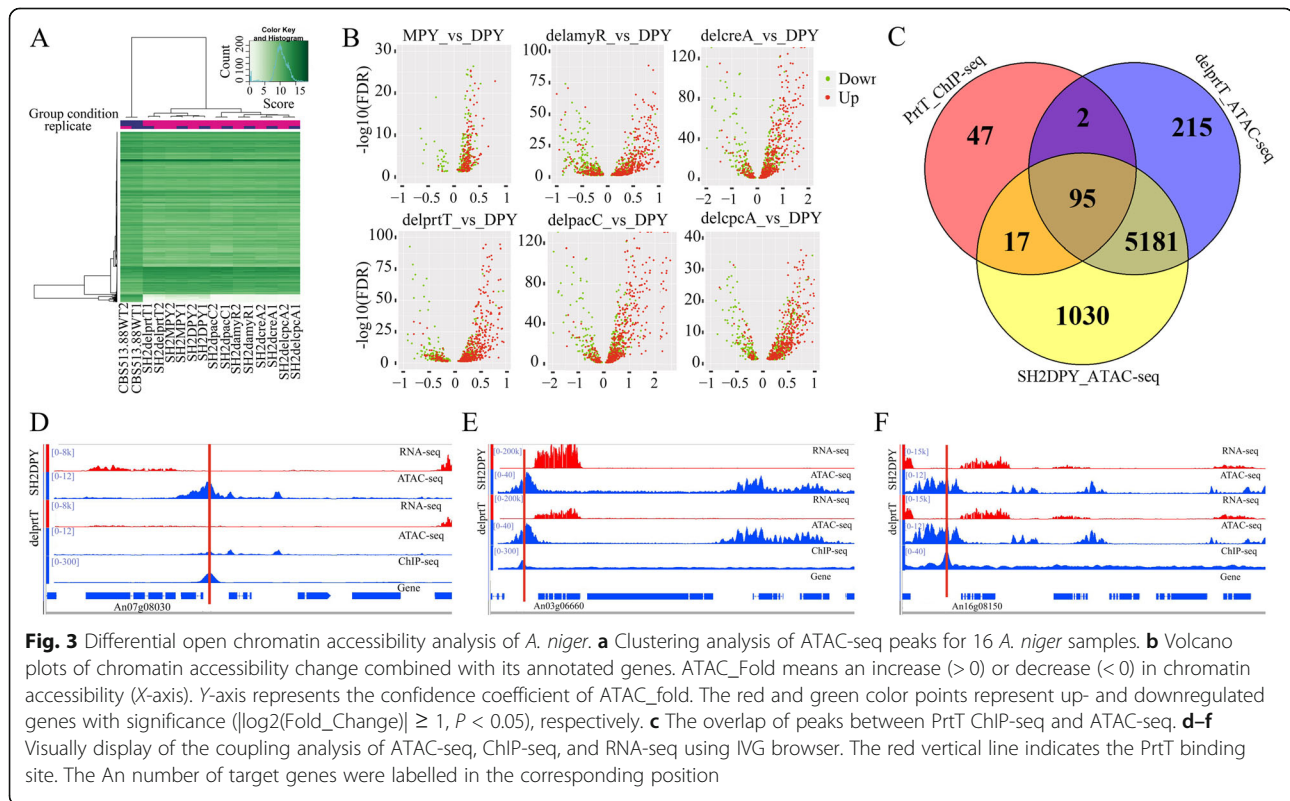
(See figure on previous page.)

**Fig. 2** Whole-genome landscape of chromatin accessibility in *A. niger*. **a** Syntenic analysis between *A. niger* CBS513.88 (contig A1-A17 corresponding to An01-An19) and SH2 (contig S1-S11), and visualization of ATAC-seq peak signals in *A. niger* CBS513.88 (orange) and SH2 (blue). The outer signal represents the *A. niger* CBS513.88 naked genomic DNA which was used as the integration control of Tn5 transposase. **b** A Venn diagram showing the overlap of accessible regions between *A. niger* CBS513.88 and SH2. **c** Density plots of mapped reads (upper) and heatmap of peak length (bottom) for the 2-kb region flanking the peak center. **d** Distribution of accessible regions in the promoter, intergenic region, exon, intron, and TTS in *A. niger* CBS513.88 and SH2. **e** Syntenic analysis of ATAC-seq peaks between *A. niger* CBS513.88 (contig A1-A17) and *A. oryzae* niaD30 (contig R1-R8). **f** Functional clustering of overlapping ATAC-seq peaks of *A. niger* CBS513.88 and *A. oryzae* niaD300

promoter regions were identified when comparing with *A. niger* SH2 cultured in DPY and MPY media (Additional file 5: Table S4). We found that changing the culture conditions could slightly alter the chromatin accessibility, in which 3853 of the DAPs were specific to *A. niger* SH2 cultured under MPY medium (Additional file 5: Table S4). Almost all upregulated genes were driven by the MPY-specific ATAC-seq peaks (Fig. 3b: MPY\_vs\_DPY). Besides, the induced conditions such as alternative carbon sources or differential pH could also slightly alter the chromatin accessibility (Data submitted to NCBI, PRJNA692847), and low-level CreA can strongly inhibit the expression of *xlnD* and *abfA* (Additional file 1: Figure S2). Furthermore, we identified TF-specific chromatin accessibility using the peak signal intensity of *A. niger* SH2 compared with that observed in TF-deleted strains (FDR < 0.05). In total, 1641 of the amyR-specific peaks, 1156 of the prtT-specific peaks,

773 of the cpcA-specific peaks, 1301 of the pacC-specific peaks, and 1305 of the creA-specific peaks in the promoter regions exhibited significantly higher accessibility (FDR < 0.05) than those in the corresponding TF-deleted strains (Additional file 5: Table S4). The distributions of upregulated genes were unbalanced and trended towards TF-specific chromatin accessibility (Fig. 3b).

To confirm the ATAC-seq data, we performed ChIP-seq of the protease regulator PrtT (NCBI accession NO.: PRJNA692789). A total of 161 specific peaks were enriched, of which 97 (60.24%) peaks could be mapped with *delprtT* and 112 (69.57%) peaks with WT (Fig. 3c). The conserved motif CCGHCGG [35] could also be enriched (Additional file 1: Figure S3). Furthermore, the coupling analysis of ChIP-seq, RNA-seq, and ATAC-seq data of PrtT (Fig. 3d–f and Additional file 1: Figure S4) showed three major regulation categories of TF: (i) the



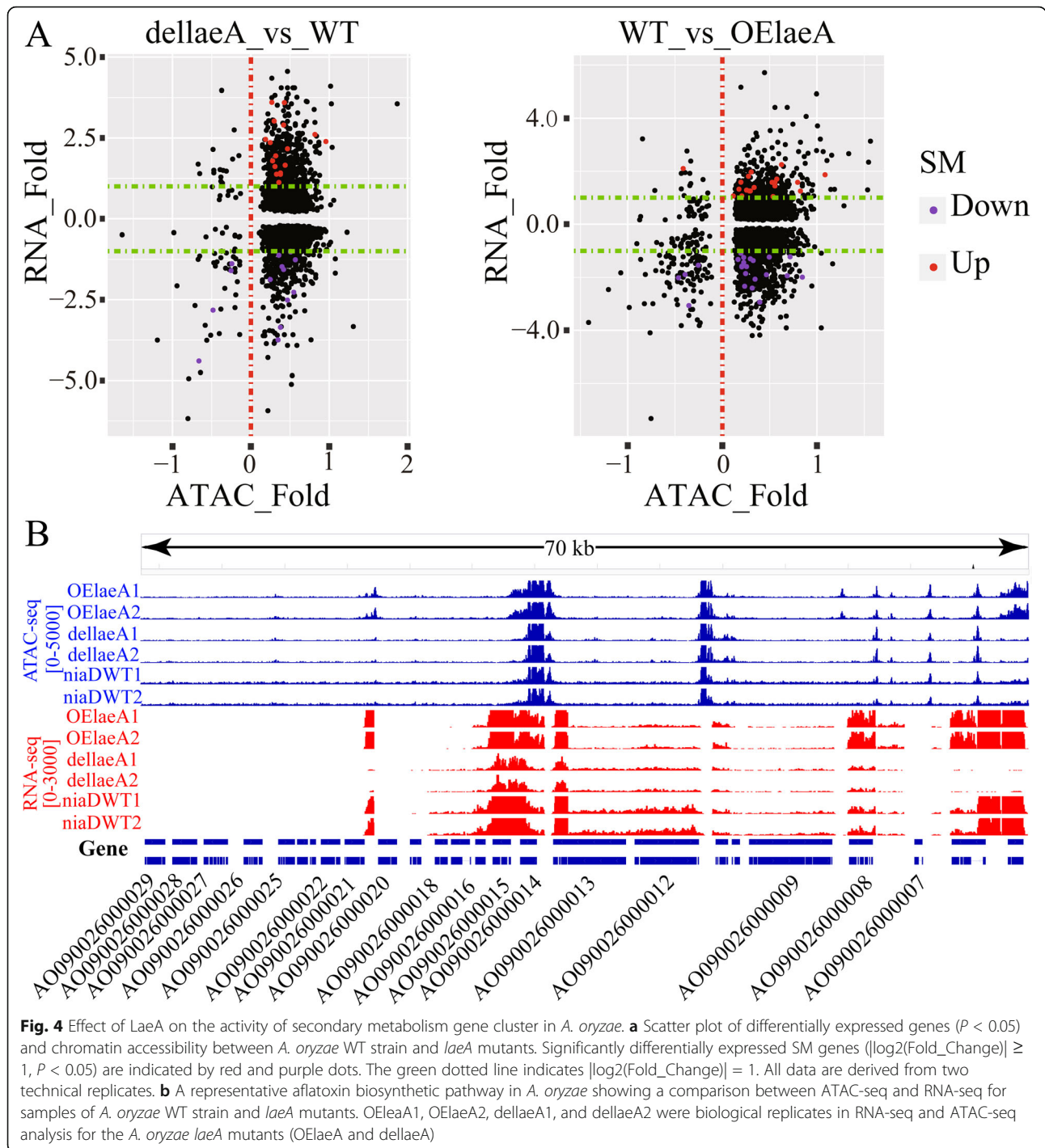
deletion of PrtT leads to decreased chromatin accessibility, thereby downregulating target genes (Fig. 3d and Additional file 1: Figure S4a); (ii) the deletion of PrtT does not affect chromatin accessibility, but downregulates target genes (Fig. 3e and Additional file 1: Figure S4b); (iii) the deletion of PrtT neither affects chromatin accessibility, nor affects gene expression (Fig. 3f and Additional file 1: Figure S4c).

The biosynthetic gene clusters of secondary metabolites (SMs) in *Aspergillus* species are frequently silent or inactive under normal laboratory conditions due to a tendency to be located in subtelomeric regions with a high degree of heterochromatin condition; however, these genes can be activated by changing the chromosomal state of the SM gene clusters from heterochromatin to euchromatin [36]. LaeA has been acknowledged as a global regulator of secondary metabolism in *Aspergilli*. Deletion of LaeA in filamentous fungi decreased the virulence of the strains and the production of secondary metabolites. And it has been reported that deletion or overexpression of LaeA could be used to explore numerous secondary metabolite clusters [37–39], although the physiological function of LaeA in modulating gene expression remains little known [40]. To investigate whether LaeA could regulate the chromatin accessibility in *Aspergilli*, we constructed ATAC-seq libraries of *dellaeA* and *OElaeA* and identified the ATAC-seq peaks with significantly changed accessibility (FDR < 0.05) in *dellaeA*-vs-WT (4725 DAPs) and *OElaeA*-vs-WT (5572 DAPs). (Figure 4a, Additional file 1: Figure S5; Additional file 2: Table S1, and Additional file 6: Table S5). SM biosynthetic gene clusters in the *A. oryzae* RIB40 genome contain 621 biosynthetic SM genes predicted by Secondary Metabolite Unique Regions Finder (SMURF) [41]. The chromatin regions in front of 138 and 164 of the SM biosynthetic genes in *A. oryzae* *OElaeA* and *dellaeA* mutants significantly changed accessibility (FDR < 0.05) to induce significant regulation of 42 and 32 SM biosynthetic genes, respectively (Fig. 4a and Additional file 6: Table S5). The 54 backbone genes of SM biosynthesis include polyketide synthases (PKSs) and non-ribosomal peptide synthetases (NRPSs) from the *A. oryzae* RIB40 genome; only 12 backbone genes of SM biosynthesis significantly changed accessibility (FDR < 0.05) to induce significant regulation of the corresponding SM backbone genes in *OElaeA* and *dellaeA* mutants, which contained 7 NRPSs, 4 PKSs, and 1 DMAT (Additional file 6: Table S5). We found that the knockout and overexpression of the *laeA* gene only up- and downregulated the SM biosynthetic genes that were expressed in the wild-type strain, while there was never de novo activation of those SM biosynthetic genes that were not expressed in the wild-type strain (Additional file 1: Figure S6). Furthermore, we investigated chromatin

accessibility and transcriptional activity in the famous aflatoxin (AF) biosynthesis pathway of *A. oryzae* (Fig. 4b). The sequences of the aflatoxin (AF) biosynthesis gene clusters were divided into two parts: active and inactive chromatin regions. The sequences from AO090026000004 to AO090026000020 were the active chromatin regions, which were characterized by highly expressed biosynthesis genes. The overexpression and knockout of the *laeA* gene were able to change the chromatin accessibility in the active regions. The key transcription factor *aflR* (AO090026000014) was in the region of high chromatin accessibility, while the coactivator *aflJ* (AO090026000015) became highly expressed following the overexpression of *laeA*. The other sequences from AO090026000020 to AO090026000031 were in inactive chromatin regions in which no ATAC-seq peaks were identified, and biosynthesis genes had no expression or the lowest level of expression. The AF biosynthesis pathway components were located in a *laeA*-dependent SM biosynthetic gene cluster. Taken together, the reason for the lack of AF biosynthesis in *A. oryzae* might be inactive chromatin regions around the AF biosynthesis pathway.

#### Identification of footprints in filamentous fungus *Aspergillus niger*

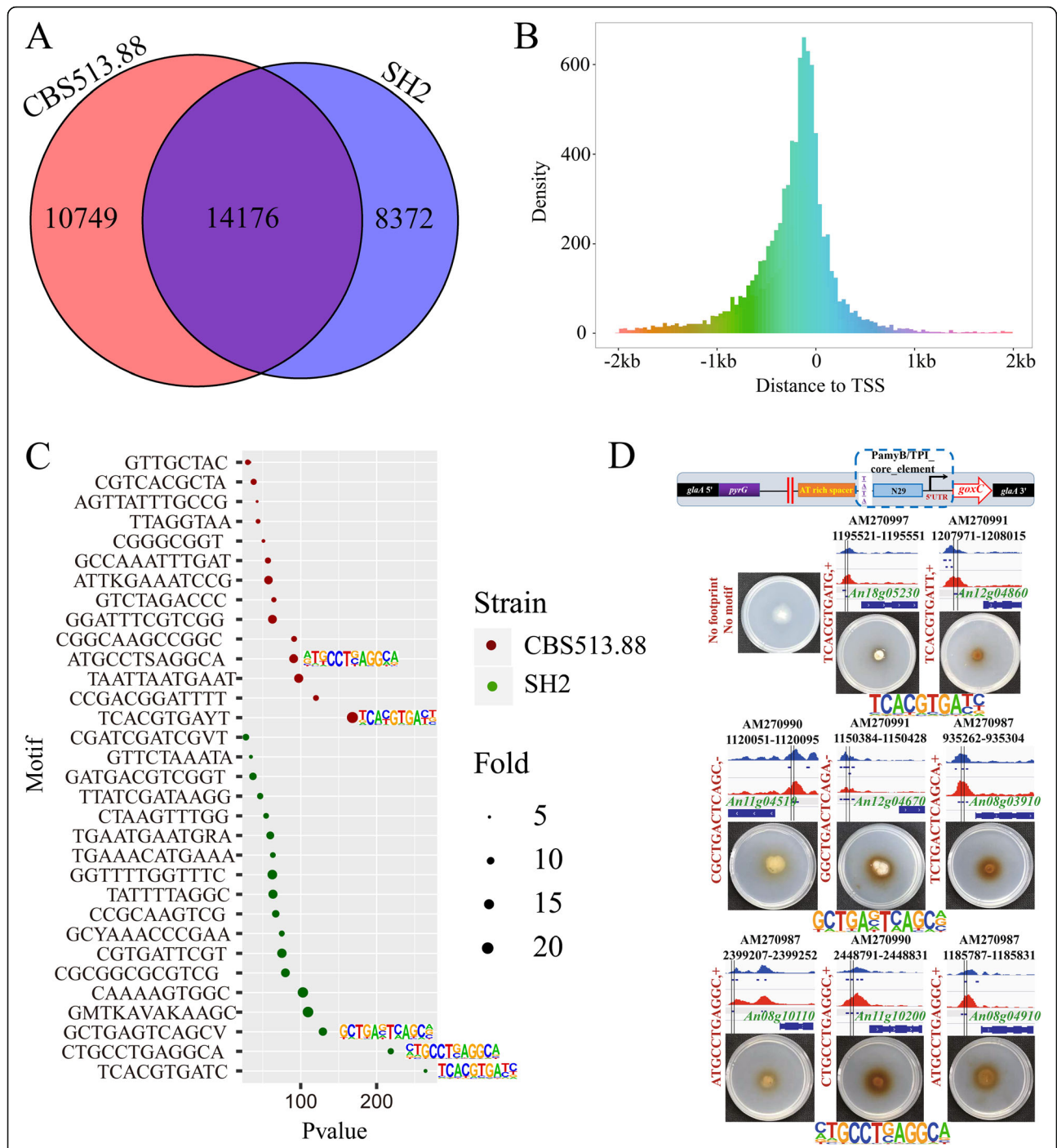
DNA regions of footprints bounded by DNA-binding proteins are protected from Tn5 integration, at which the read coverage suddenly drops [26, 42]. A high-depth ATAC-seq can be used to identify depleted narrow regions in open chromatin regions of a genome corresponding to a single TF footprint [26, 42, 43]. We used the “*wellington\_footprints.py*” function of the pyDNase package [44] to systematically identify Tn5 integration-insensitive sites from CBS513.88 and SH2 and footprints from ATAC-seq peaks (Fig. 5a). In total, there were 24,925 and 22,548 DNA footprints detected in vivo across the *A. niger* genome from 7297 and 7697 ATAC-seq peaks of CBS513.88 and SH2, respectively (footprint data from all wild-type and TF-deleted strains can be found in Additional file 7: Table S6). A total of 14,176 DNA footprints overlapped between CBS513.88 and SH2 ATAC-seq data. The DNA footprints identified by the ATAC-seq data were enriched 1-kb upstream and 0.5-kb downstream of the TSSs (Fig. 5b). To further characterize genomic footprints, we performed de novo motif enrichment analysis on the footprints from CBS513.88 and SH2 using the “*findMotifsGenome.pl*” function of the HOMER package with the default settings [45]. A total of 15 and 17 de novo TFBS with a *P* value threshold of  $1.0E-12$  were enriched more than 5-fold from the two sets of footprint data from CBS513.88 and SH2, respectively (Fig. 5c and Additional file 8: Table S7). De novo TFBS of other SH2 TF-deleted



strains with a  $P$  value threshold of  $1E-50$  were also identified using the HOMER method (Additional file 1: Figure S7 and Additional file 8: Table S7). The E-box of bHLH factors (motif TCACGTGATC), the SltA homolog binding motif (motif CTGCCTGAGGCA), and the core element of bZIP factors (motif GCTGAGTCAGCV) were among the top 5 most enriched motifs according to  $P$  value in all de novo TF binding motifs (Fig. 5c).

Instances of each TF binding motif were identified using the “*annotatePeaks.pl -m*” function of the HOMER package to search for DNA sequences in the 24,925 DNA footprints from *A. niger* CBS513.88 and the 22,548 DNA footprints from *A. niger* SH2 (Additional file 9: Table S8). Instances of genes controlled by de novo TF binding at motifs were identified in footprints in the promoter regions, and they were analyzed for the



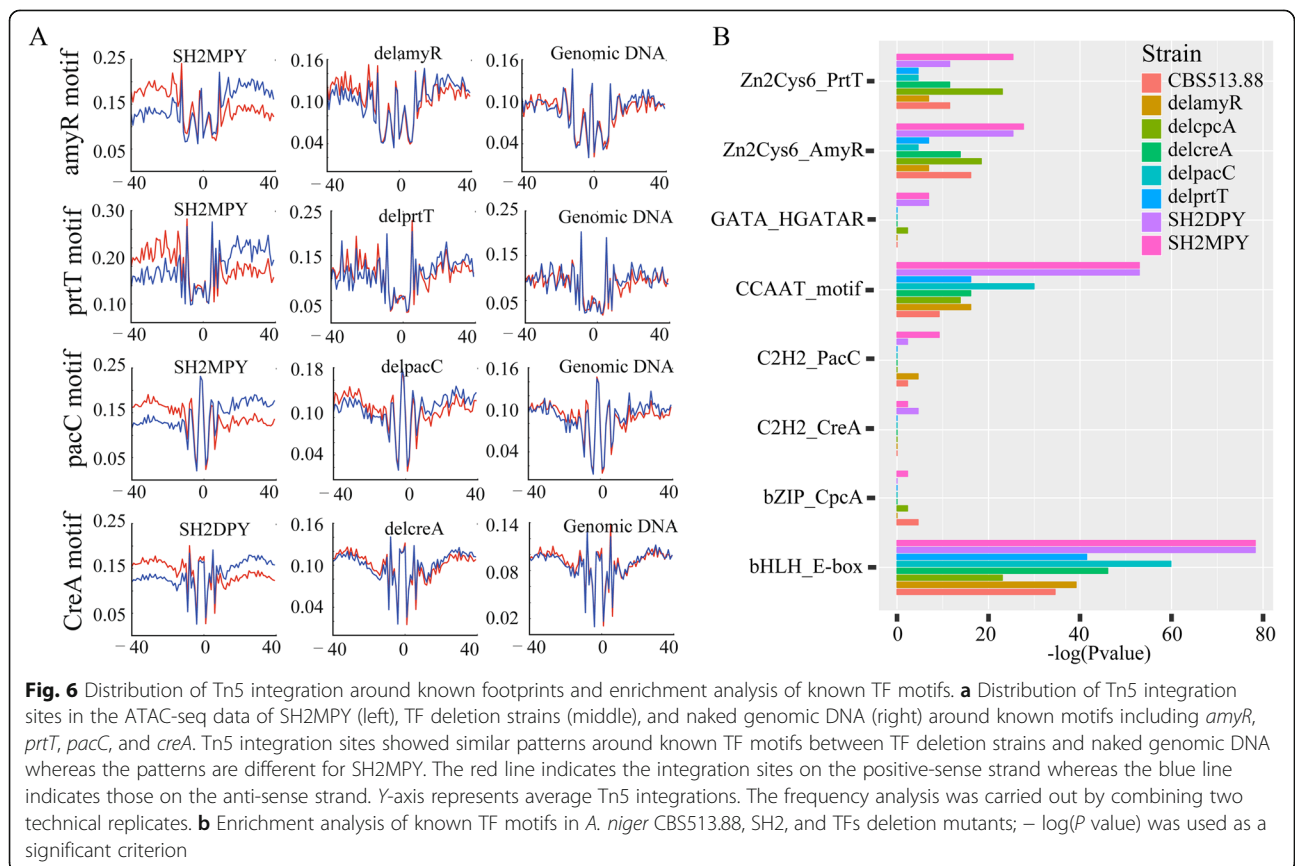


**Fig. 5** Identification of de novo predicted footprints in *A. niger* CBS513.88 and SH2, and functional verification in regulating gene expression. **a** Venn diagram analysis of identified footprints in *A. niger* CBS513.88 and SH2. **b** Statistical analysis of the distances between the identified DNA footprints and the TSSs. **c** De novo prediction of transcription factors in *A. niger* CBS513.88 (red) and SH2 (green). The circle size represents the enrichment fold. The top three transcription factor motifs in *A. niger* SH2 are labeled with the corresponding binding motifs. The x-axis represents *P* value. **d** In vivo functional verification of putative de novo transcription factor targeting sites were carried out by artificially synthesized minimal promoters. The red dividing line represents the location of the footprints to be verified. The red text on the side is the TF binding motif contained in the verified DNA instances. From top to bottom, there were 2 E-box of bHLH factors (motif TCACGTGATC), 3 core elements of bZIP factors and 3 SitA homologs binding motif (motif CTGCCTGAGGCA) target sites validated by phenotype plates in vivo

functional enrichment of corresponding TF binding motifs (Additional file 1: Figure S8). In the *A. niger* SH2, 269 genes under the control of the de novo motif TCACGTGATC (E-box) were enriched in GO categories “structural constituent of the ribosome” and “translation” (Additional file 1: Figure S8a). In addition, 154 genes under the control of the de novo motif CTGCCTGAGGCA (SltA homolog) were enriched in the molecular function GO categories “ATP binding” and “structural constituent of the ribosome;” in the cellular components of “ribosome,” “mitochondrial inner membrane,” and “vacuole;” and for the biological processes of “translation” and “proteolysis” (Additional file 1: Figure S8b). Then, 54 genes under the control of the de novo motif GCTGAGTCAGCV (bZIP factors) were also enriched in the molecular function GO categories “zinc ion binding,” “DNA binding,” and “transcription factor activity” in the nucleus for the process of regulation of transcription (Additional file 1: Figure S8c).

Furthermore, to detect whether the ATAC-seq footprints containing the TF binding motifs drove the expression of functional target genes, we constructed a minimal synthetic promoter consisting of a core element, an ATAT-binding box, an AT-rich spacer, and a footprint library-driven UAS element with a TF binding site motif (Additional file 1: Figure S9) [46]. A 29-bp

sequence from the TSS in the *amyB* (An05g02100) promoter of *A. niger* was chosen as the core element due to the high-level expression of the *amyB* gene. The glucose oxidase *goxC* (An01g14740) was used as a reporter, the expression of which would be revealed by a colored colony. The repression under ER stress (RESS) element identified in our prior study [46] was used as a positive control, and it also produces colored colonies (Additional file 1: Figure S9). As a negative control, a sequence containing only the *amyB* core element, the ATAT sequence, and the AT-rich spacer was constructed (no colored colonies but basal transcription can occur (Additional file 1: Figure S9)). We used this minimal promoter system to determine whether the ATAC-seq footprints containing de novo motifs, such as motif TCACGTGATC (E-box), motif CTGCCTGAGGCA (SltA homolog), and motif GCTGAGTCAGCV (bZIP factors), could drive the expression of target genes (Fig. 5d). Verifying the translation and ribosome functions controlled by the motif TCACGTGATC (E-box), the footprint of An12g04860, which encodes a ribosomal protein of the large subunit L30, was found to drive the expression of the reporter gene in vivo, resulting in colored colonies (Fig. 5d). The footprints containing motif CTGCCTGAGGCA located in the promoter regions of An08g04910 and An11g10200, which encode the



NADH-ubiquinone oxidoreductase and cytochrome C oxidase subunit in the electron respiratory chain, respectively, also drove the expression of the reporter gene in vivo. The footprints containing the motif GCTGAG TCAGCV, which is located in the promoter regions of An11g04510 and An12g04670 (related to translation release factor activity and translation initiation factor, respectively), were confirmed to drive the expression of the reporter gene in vivo. Taken together, genome-wide DNA footprints were systematically identified from accessible chromatin regions based on our ATAC-seq data. In addition, de novo motifs enriched from ATAC-seq footprints were able to drive the expression of their target genes in vivo.

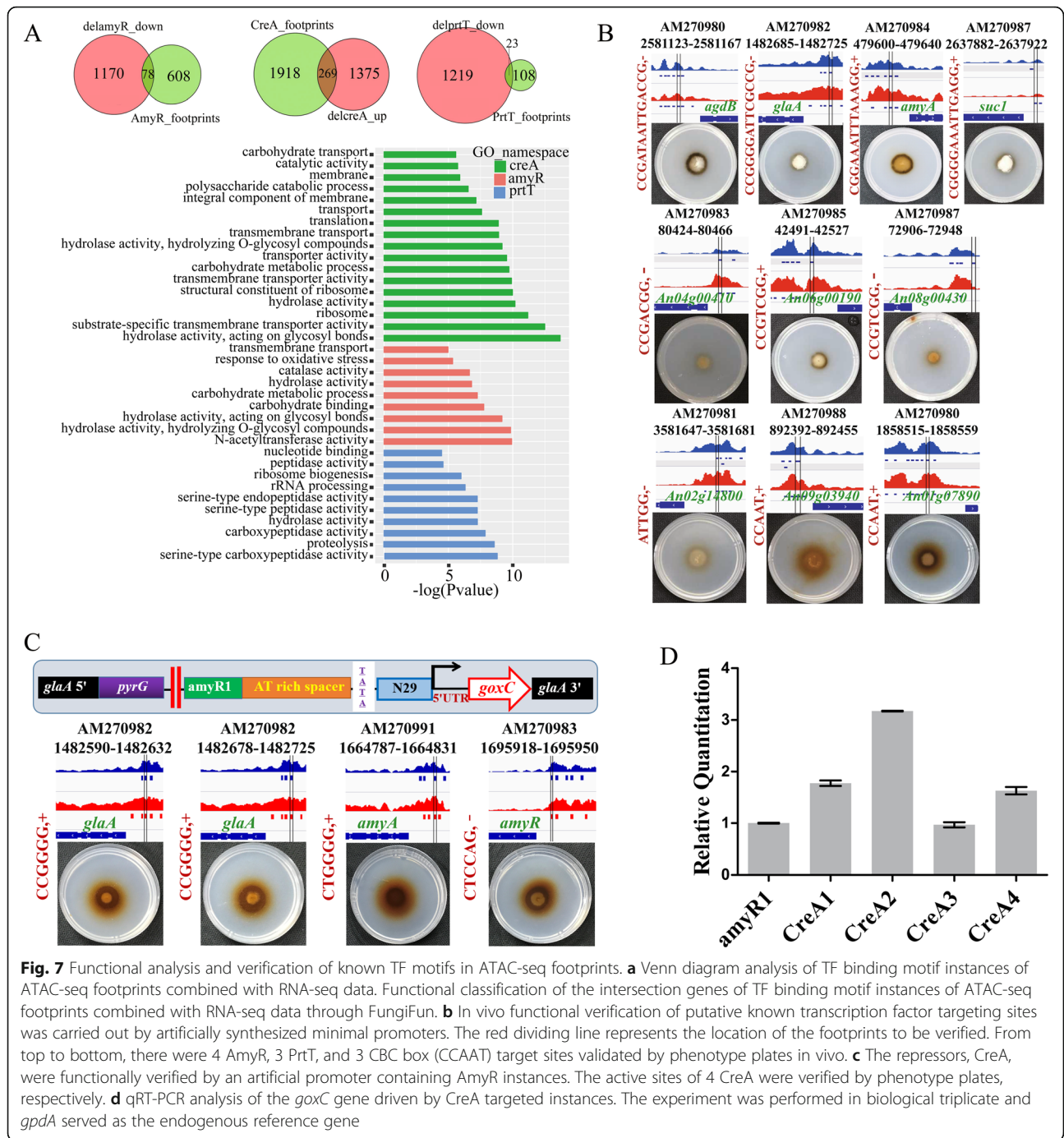
#### Known TF motifs of *Aspergillus niger* enrichment in ATAC-seq footprints

We determined the transposase Tn5 integration patterns around the binding sites of AmyR, PrtT, PacC, and CreA in the *A. niger* SH2 cultured under MPY medium (Fig. 6a), while these patterns for known TFs were also evident in naked genomic DNA (Fig. 6a). We further found that distributions of Tn5 integration sites in the flanking regions of binding site motif sequences were asymmetrical when comparing the anti-sense and sense strand and the patterns which were completely different from the symmetrical patterns observed in the naked genomic DNA. Furthermore, to elucidate whether strand-specific Tn5 integration patterns around TF binding sites are caused by protein/DNA interactions, we calculated the Tn5 integration frequency in TF-deleted strains (Fig. 6a). In this experiment, the conserved DNA motif should not be enriched, for there was no corresponding known TF to be expressed in vivo. Tn5 integration sites around the known TF binding sites in the absence of the corresponding TF condition showed a similar pattern to that of genomic DNA, which was a symmetrical integration pattern around TF binding sites on both DNA strands. The asymmetrical pattern of Tn5 integration around TF binding sites in our ATAC-seq data was distinctly different from the no TF binding pattern on both TF-deleted strains in vivo and naked genomic DNA in vitro (Fig. 6a). We demonstrated here that strand-specific Tn5 integration patterns around TF binding sites are caused by protein/DNA interactions, which could be used as a signal indicating an active TFs.

To determine the enrichment of known TF binding motifs from ATAC-seq footprints by comparing ATAC-seq data from *A. niger* WT strain and its TF-deleted strains, we collected the available binding motif sequences of 19 known *Aspergillus* TFs [23] and employed the “findMotifsGenome.pl” function of the HOMER package to identify overrepresented motifs of known *Aspergillus* TFs from ATAC-seq footprint data. We found

that the binding motifs of bZIP CpcA, C2H2 CreA, and PacC were not overrepresented in the ATAC-seq footprints of the corresponding TF-deleted strains ( $P$  value  $> 1.0E-2$ ) (Fig. 6b). Although the binding motifs of the Zn<sub>2</sub>Cys<sub>6</sub> family AmyR and PrtT were also enriched in the footprints from the TF-deleted strains delamyR and delprtT, respectively, the enrichment  $P$  values of AmyR and PrtT motifs were less than  $1.0E-2$  (Fig. 6b). The binding motifs for bHLH family E-box, CCAAT binding complex (CBC), Zn<sub>2</sub>Cys<sub>6</sub> family AmyR, and PrtT were significantly enriched in all *A. niger* strains, especially the overrepresentative enrichment of SH2 in MPY medium (Fig. 6b). Other TF family motifs, such as bZip CpcA, GATA AreA, C2H2 CreA, and PacC, were only enriched in the ATAC-seq footprints of the SH2 cultured in MPY and DPY medium (Fig. 6b). Footprints that contained a significant motif match were considered to be instances of predicted binding motifs. We used the “annotatePeaks.pl -m” function of the HOMER package to detect the motif instances of known TFs from footprints by comparing *A. niger* WT strains and the corresponding TF-deleted strains (Additional file 10: Table S9). The motifs of AmyR, PrtT, PacC, and CreA were identified in 765, 158, 821, and 2865 motif instances in the SH2 cultured in MPY, respectively (Additional file 10: Table S9). However, we still detected 593 instances of AmyR binding, 103 instances of PrtT binding, 656 instances of PacC binding, and 2648 instances of CreA binding from footprints in the TF-deleted strains delamyR, delprtT, delpacC, and delcreA, respectively (Additional file 10: Table S9).

Furthermore, to investigate whether instances of TF binding drove the expression of target genes, we used TF binding motif sites in ATAC-seq footprints combined with RNA-seq data for the changes in gene expression between the SH2 cultured in MPY and the corresponding TF-deleted strains (Fig. 7a). The expression of 1248 genes in the delamyR and 1242 genes in the delprtT was downregulated compared to what was seen in the control of the SH2 strain (Fig. 7a). Only 78 and 23 genes were downregulated in the delamyR and delprtT strains overlapped with motif target genes, respectively (Fig. 7a). As expected, 23 target genes under the control of PrtT were mainly involved in serine-type peptidase activity, and 78 target genes under the control of AmyR were functionally enriched in hydrolase activity, acting on glycosyl bonds (Fig. 7a). A total of 1644 genes were upregulated in the delcreA compared with the control of the SH2 strain, in which 269 genes overlapped with motif target genes from CreA (Fig. 7a). GO functional enrichment analysis revealed that 269 genes under the control of CreA were primarily involved in the molecular function of hydrolase activity and transporter activity. Besides, the enriched genes were



**Fig. 7** Functional analysis and verification of known TF motifs in ATAC-seq footprints. **a** Venn diagram analysis of TF binding motif instances of ATAC-seq footprints combined with RNA-seq data. Functional classification of the intersection genes of TF binding motif instances of ATAC-seq footprints combined with RNA-seq data through FungiFun. **b** In vivo functional verification of putative known transcription factor targeting sites was carried out by artificially synthesized minimal promoters. The red dividing line represents the location of the footprints to be verified. From top to bottom, there were 4 AmyR, 3 PrtT, and 3 CBC box (CCAAT) target sites validated by phenotype plates in vivo. **c** The repressors, CreA, were functionally verified by an artificial promoter containing AmyR instances. The active sites of 4 CreA were verified by phenotype plates, respectively. **d** qRT-PCR analysis of the *goxC* gene driven by CreA targeted instances. The experiment was performed in biological triplicate and *gpdA* served as the endogenous reference gene

clustered in the structural constituents of the ribosome, the cellular components of membranes and ribosomes, and their biological processes involved in carbohydrate catabolism, carbohydrate transport, and translation (Fig. 7a).

To experimentally assess whether footprints with these known TF binding motifs that were identified by ATAC-seq data were functional TF binding sites, we performed footprint-driven reporter assays in vivo (Fig. 7b). Three

footprints containing the AmyR binding motif were located in promoters of glycosyl bond hydrolase genes: An01g10930 (*agdB*), An03g06550 (*glaA*), and An05g02100 (*amyA*). These AmyR binding motifs drove the expression of the reporter gene, resulting in colored colonies. Three footprints containing the PrtT binding motif in the promoter of peptidase genes, including An04g00410 (dipeptidyl peptidase III), An06g00190 (Sedolisin family secreted protease), and An08g00430

(serine-type carboxypeptidase), also drove the expression of the reporter gene. In addition, footprints containing CCAAT box or AreA binding motifs can also function as transcription activator (Fig. 7b and Additional file 1: Figure S10).

In addition to identifying the activation function of ATAC-seq footprints containing the known TF binding motifs, we further investigated whether ATAC-seq footprints containing repressor binding motifs could down-regulate the expression of target genes (Fig. 7c). We constructed a minimal synthetic promoter with the footprint-driven UAS elements containing the AmyR binding motif instance (An01g10930, *agdB*) as a positive control (Fig. 7c). The footprint containing the repressor binding motif was assessed by assembling downstream of the AmyR footprint. *A. niger* CreA (An02g03830, contains two Cys<sub>2</sub>His<sub>2</sub> zinc finger motifs) is the transcriptional regulator that mediates carbon catabolite repression [47]. However, our results showed that footprints containing CreA binding motif sites located in promoters of glycosyl bond hydrolases such as *glaA* (An03g06550), *amyA* (An12g06930), and *amyR* (An04g06910) were able to upregulate the expression of the reporter gene (Fig. 7c, d). Two footprints containing the CreA binding motif in the *glaA* promoter possessed the same ability to activate functional gene expression (Fig. 7c). Taken together, the transcriptional regulator CreA acts not only as repressor but also as activator.

## Discussion

In ATAC-seq, the Tn5 transposase integrates its adaptor payload into locations of open chromatin, which can be amplified by PCR and assessed with high-throughput sequencing to study protein/DNA interactions on a genome scale [27]. Here, we developed an ATAC-seq protocol for filamentous fungi to identify chromatin accessibility on a genome scale. Our initial attempts failed to generate ATAC-seq data from the chromatin following digestion of Tn5 transposase, harvesting the mycelium and reducing it to powder by grinding it in liquid nitrogen with a mortar and pestle. The gel image analysis shows no discrete nucleic acid bands, and this method resulted in a high noise-to-signal ratio after sequencing. In our modified *Aspergillus* ATAC-seq protocol, the lysis of the *Aspergilli* cell wall was carried out by adding hydrolase during cultivation, and protoplasts were obtained by filtration through a miracloth membrane. High-quality intact nuclei released from *A. niger* protoplasts, which are similar to mammalian cells, were used to improve the data collected from ATAC-seq. ATAC-seq requires few cells, so a sufficient number of protoplasts were easily obtained from *A. niger* cultured under various conditions. Yeast spheroplasts from *S. cerevisiae* and *Schizosaccharomyces pombe* have also been

used in ATAC-seq assays [33]. Furthermore, protoplast preparation could occur with *A. niger* cells cultured under various conditions to decrease experimental variation. In our *A. niger* ATAC-seq library, we detected the same periodicity pattern in the insert size distribution of sequenced fragments from chromatin as what was observed in a human lymphoblastoid cell line [27] and in *Arabidopsis thaliana* [43]. The *A. niger* ATAC-seq library has a high signal-to-noise ratio, which was controlled by comparing the data to Tn5 integration into naked genomic DNA. The reproducibility of biological and technical replicates in our *A. niger* ATAC-seq protocol was relatively high. In ATAC-seq experiments using mammalian cells and plant cells, typically ~30–70% of the sequenced reads are aligned to the organellar genomes [27, 43]. In our *A. niger* CBS513.88 ATAC-seq results, a significantly higher fraction of reads were mapped to the nuclear genome. These data show that our *Aspergillus* ATAC-seq protocol is an effective method that can be applied to the analysis of other species of filamentous fungi.

The chromatin accessibility regions of *A. niger* CBS513.88 and SH2 were similarly distributed across the various elements of the genome and showed a high proportion at TSSs (~70%) and a low proportion at intergenic regions (~5%), which was contrary to humans [48], *Arabidopsis* [32], and tomato [8]. This difference may reflect the smaller size of the *A. niger* genome (35 Mb) [3] compared with the larger and less compact genomes of humans and plants. In addition, TTS also contained a small number of accessibility regions (~13%), which is similar to what was observed in the mouse [28]. Therefore, we reasoned that TTSs may also play an important role in the transcriptional regulation of genes and downstream translational control in *A. niger*.

Prior works found that the accessible chromatin regions were positively correlated with gene expression changes of the nearest target genes [28, 32, 49]. However, the regulatory regions could also be suppressed by some TFs or nucleosomes [49]. Our results further verified the coexistence of transcriptional activation and inhibition in the accessible region. Besides, the gene transcription activation can be accomplished by a single TF alone, or by a combination of multiple TFs (Fig. 3). Under the condition of single activation, TF-deleted directly affects chromatin accessibility (Fig. 3d and Additional file 1: Figure S5a). For multi-gene regulatory regions, there is a competitive regulation of TFs [50, 51]. When TF is missing, other proteins occupy the vacated sites, thereby maintaining chromatin accessibility. In this case, the transcription of downstream gene is jointly determined by the lack of TF and the newly occupied TFs.

Comparative genomics focuses on the comparison of DNA coding regions between different species, revealing

the eukaryotic genomic evolution of different *Aspergillus* species [9]. In comparing open chromatin regions (non-coding sequences) between *A. niger* and *A. oryzae* species, we discovered that syntenic orthologues of 1564 accessible regions across *A. niger* and *A. oryzae* drive the expression of a collection of target genes that are significantly enriched in housekeeping genes functioning in ribosomes, translation, and energy. The synteny of open chromatin regions between different species may be due to the conservation of noncoding regions [9], indicating that *Aspergillus* species, which act as protein expression cell factories, exhibit the same potential functional elements in protein synthesis regulation.

The accessible chromatin regions identified by ATAC-seq are particularly valuable for the identification of cis-regulatory DNA elements [32, 48, 49]. By comparing the Tn5 digestion patterns of WT, TF-deleted strains, and naked genome, the strand-specific Tn5 integration patterns around TFs were only observed in wild-type strain (Fig. 6a). In fact, TF does not exist in the naked genomic DNA and the TF-deleted strain, so the DNA footprints should not be enriched. Given the digestion bias of Tn5 transposase, these footprints are regarded as the pseudo footprints and several research proposed correction methods of TF footprinting [26, 42, 52]. However, there is no consensus on the significance of correcting DNase I or Tn5 transposase sequence bias. Li et al. [26] developed HINT-ATAC to measure the amount of the strand cleavage bias for distinct fragment sizes (All, Nfr, 1 N, and + 2 N) around distinct intervals near the TF binding site, and the results showed intricate strand-specific cleavage patterns relative to TF binding. Therefore, we reasoned that the strand-specific unbalanced pattern of Tn5 digestion on both sides of the footprints can be considered a true DNA footprint.

We systematically identified genome-wide DNA footprints from candidate accessible regions and performed de novo motif enrichment analysis on the footprints from *A. niger* CBS513.88 and SH2 (Fig. 5a, c). Clustering of TFs by family and analysis of *P* values indicated that the *A. niger* CBS513.88 and SH2 ATAC-seq footprints were enriched for de novo bZIP (motif GCTGAGTCAG CV) and basic helix-loop-helix (motif TCACGTGATC) families (Fig. 5c). The 775 genes predicted as DNA-binding TFs are primarily distributed among 37 classes of regulatory proteins, including 17 bZip family TFs and 10 bHLH family TFs in the *A. niger* genome [3, 5].

The bZip family of TFs that have been functionally analyzed in *A. niger* mainly include *hacA* (UPR) and *cpcA* (amino acid starvation). The function of other bZip TFs are associated with amino acid starvation (*jlba*); oxidative stress (*napA*, *atfA*, and *atfB*); and sulfur metabolism (*cys-3/metR*), and they have mainly been characterized in homologous species, such as *A. nidulans* [53–55], *A.*

*fumigatus* [14], and *A. oryzae* [56, 57]. The gene regulatory network of the bZIP family of *Neurospora crassa* [58] reveals that the bZIP family of TFs exerts consistent regulatory effects in response to environmental stresses, such as oxidative stress, amino acid starvation, and heavy metal pressure. Functional enrichment of the most proximal gene targeted by the bZip motif, GCTGAGTCAG CV, further demonstrates that this type of TF is associated with transcriptional regulation, which is the first step in regulating the stress response of filamentous fungi to the external environment.

The 10 TFs in the bHLH family, including homologous genes of *sclR/srbB*, *anbH1*, *srbA*, *palcA*, *INO4*, and *devR*, have not been well characterized in *A. niger*. However, in other species, the bHLH TFs have been postulated to be involved in hyphae growth and differentiation as well as specific metabolic pathways [36, 59]. The GO category of our E-box targeted genes found that they are involved in the biological process of translation, which is identical to the function of the bHLH family proteins that was predicted following *A. nidulans* genome sequence analysis [9]. Thus, we reasoned that bHLH TFs in *A. niger* have more unknown functions, and further characterization of bHLH TFs is urgently needed. The bHLH TFs recognize a consensus core element named the E-box (5'-CAVVTG-3'), and the G-box (5'-CACGTG-3') is the most common form [59]; further, specific recognition of the sequence relies on flanking sequences [60, 61]. The E-box flanking sequences predicted herein contain not only the 5' T residue but also the 3' ATC residues. The 5' T residue flanking the G-box determines major specificity by preventing the binding of other bHLH TFs, such as yeast PHO4 [61], whereas the significance of the 3' residues is not fully known. The new motif CTGCCTGAGGCA is also a highly enriched motif according to the *P* value compared to all de novo previously unidentified TF binding motifs. GO functional analysis showed that this new motif CTGCCTGAGGCA drives the expression of target genes, and the GO molecular function categories include the cellular components of "ribosome" and "mitochondrial inner membrane;" the biological processes are related to "translation" and "metabolism."

We constructed a minimal synthetic promoter suitable for filamentous fungi that was based on the yeast minimal promoter [62]. We used this promoter to validate the footprint library-driven UAS elements with TF binding site motifs, which provided a new method for in vivo validation of *A. niger* ATAC-seq-predicted binding sites containing TFs in the absence of other omics data (DNase-seq, ChIP-seq). Our de novo motifs identified from enriched from ATAC-seq footprints were able to drive the expression of the targeted gene in vivo.

We further determined the enrichment of known TF binding motifs from ATAC-seq footprints by comparing ATAC-seq data from the *A. niger* wild-type strain and its TF-deleted strains. Footprints for these known TF binding motifs identified by ATAC-seq data were also verified as functional TFs in vivo. Compared with *A. niger* CBS513.88, the enrichment *P* value of the TF motif in *A. niger* SH2 was significantly higher; the most differential values were observed for the CCAAT box, which represents a gene transcriptional activation domain [23], and the translation-related bHLH family [9]. This suggests that *A. niger* SH2 is a more suitable host for exogenous protein expression [4] at the TF level. The TF motif enrichment observed after growth in different carbon source conditions and in comparison to TF-deleted strains offer valuable information about TF dynamic responses in addition to the external environment and/or intrinsic regulatory factors. Transcriptome analysis of the known TFs *creA*, *amyR*, *pacC*, and *prtT* revealed that these transcription factors are involved in decentralized functions and do not cluster in categories directly related to TF function; this is even true for the pathway-specific transcription factor PrtT, which has a variety of functions that are not directly related to TF function that is clustered. Such results are ubiquitous in other studies using RNA-seq to study the function of transcription factors [13, 50, 63], which makes it particularly difficult to confirm the specific regulatory functions of transcription factors; as a result, the targeted GO term of a TF can only be determined by combining phenotypic changes [50] or ChIP-seq analysis [12, 13]. These functional clustering results are likely to be the effects of cascades of amplified changes caused by TF defects but are not the direct result of TF/DNA interaction. Through the co-regulation analysis of ATAC-seq and RNA-seq of SH2 TF-deleted strains (Fig. 7a), we effectively identified gene sets related to the corresponding TFs, and using GO clustering analysis, we screened genes and revealed that the obtained clustering results can accurately reflect the function of the experimentally verified TF (Fig. 7a) [35, 47, 64, 65]. Interestingly, we still predicted a different number of corresponding TF binding motifs in the TF knockout strains (delamyR, delprtT, delcreA, delpacC). This result may be caused by three factors: (i) None of these TFs are pioneers that change chromatin accessibility and the vacancies caused by TF defects are filled by other non-functional proteins. (ii) TF regulation is a complex dynamic process in vivo [8, 28], so that even if a TF was knocked out, there may be different TFs with alternate binding at the same DNA site. For example, the AmyR recognizing motif 5'-CGGN8(C/A)GG-3' and the PrtT recognizing motif CCG(H)CGG compete for the binding sites in the same target gene during nitrogen and carbon source

metabolism regulation [50]. (iii) These binding site identifications may be the result of a false positive. Unexpectedly, we discovered that CreA acts not only as repressors but also as activators, which was confirmed by footprint-driven reporter assays in vivo (Fig. 7c, d). Our results contradict the fact that CreA is the repressor for CCR. It can only be speculated that CreA may play an unknown activation role in gene transcriptional regulation or that the predicted binding sites contain an activated binding motif for other unknown transcription factors.

## Conclusions

These data representing the ATAC-seq landscape of TF footprints will help in the exploration of genome-wide regulation of genes active in filamentous fungi, and the data will help to discover new regulatory factors and their potential sites of action. We demonstrated the usefulness of this resource by testing 25 noncoding DNA sites with predicted TF binding motifs driving reporter gene expression in the host strain. Future experiments will be required to validate the functional activity of more noncoding sites and investigate the role of specific TF footprints in controlling gene expression. The sites with *cis*-acting elements provide a large number of basic modules for artificially synthesized promoters, and our ATAC-seq data set will provide a rich resource for the identification of functional TFs in *A. niger*. In the future, ATAC-seq may be effectively applied to study other filamentous fungi whose genomes and transcriptomes are available but whose regulatory mechanisms have not yet been explored.

## Methods

### Strains and media

The specific genotypes of strains used in this study are shown in Additional file 11: Table S10. *E. coli* Mach 1 T1 (Invitrogen, America) was used for molecular cloning and was cultured in Luria-Bertani (LB) medium (1% NaCl, 1% tryptone, and 0.5% yeast extract) with 100 µg/mL ampicillin at 37 °C (both solid and liquid) while shaking at 200 rpm. Conidial *Aspergillus niger* CBS513.88 and *Aspergillus oryzae* niaD300 were cultured in PDA medium for spore germination, and they were cultured in DPY (also called YPD) medium (2% glucose, 1% peptone, 0.5% yeast extract, 0.5% KH<sub>2</sub>PO<sub>4</sub> and 0.05% MgSO<sub>4</sub>·7H<sub>2</sub>O) for protoplast preparation. Aconidial *Aspergillus niger* SH2 ( $\Delta ku\Delta pyrG$ ) was used as the host for the knockout of transcription factors and was cultured in modified liquid Czapek-Dox medium (CD) containing the following substances to support mycelial growth: 2% glucose, 0.3% NaNO<sub>3</sub>, 0.1% KH<sub>2</sub>PO<sub>4</sub>, 0.05% MgSO<sub>4</sub>·7H<sub>2</sub>O, 0.2% KCl, and 0.01% FeSO<sub>4</sub>·7H<sub>2</sub>O. When the OD reached 2.0, 1 ml of mycelia was

inoculated into 100 ml of DPY medium and grown in a thermostatic shaker at 250 rpm and 30 °C for 40 h. Mycelia were harvested for protoplast preparation. To study the influence of carbon sources on the chromatin accessibility, the glucose in DPY was replaced with 2% maltose (MPY) [66] or strains were cultured in 2% fructose for 2 h, then transferred to CD with 0.1% fructose or CD with 1% glucose for CCR analysis [67]. *A. niger* HL-1 [68] was used as the expression host to produce the enzyme product of the glucose oxidase gene *goxC* (An01g14740). Solid CD (2% agarose) with 1% O-dianisidine (dissolved in methanol), 18% glucose, and 90 U/mL horseradish peroxidase was used to assess the color reaction following *goxC* expression.

#### ATAC-seq library preparation

The *A. niger* strain CBS513.88, SH2 (including TF deletion strains and different culture conditions) and *A. oryzae* niaD300 (including *laeA* deletion and overexpression mutants) were grown to mid-log phase in DPY medium or conditional induction medium. Then, 1 g of mycelia per library was harvested using miracloth, and they were then rinsed with cold water and 0.8 M NaCl. Protoplasts were prepared by incubating mycelial samples in enzymatic lysis buffer composed of 0.8 M NaCl, 2% cellulase, 1% helicase, 1% lyticase, and 0.5% lysozyme, which were dissolved in the corresponding medium and were then filtered with miracloth and washed with cold STC buffer (10 mM Tris-HCl pH = 7.5, 1.2 M sorbitol, 50 mM CaCl<sub>2</sub>). Subsequently, 5 × 10<sup>4</sup> protoplasts (cells) were incubated with 200 µl of lysis buffer (10 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3 mM MgCl<sub>2</sub>, 0.05% (v/v) IGEPAL CA-630, 1 mM PMSF and 1xPIC (EDTA free)) for 10 min at 4 °C, washed with lysis buffer (without IGEPAL CA-630), and then incubated with 5 µl of Tn5 Transposase of TTE Mix V50 in 45 µL of 1× TTBL buffer (TruePrep<sup>®</sup> DNA Library Prep kit V2 for Illumina<sup>®</sup>, Vazyme, China) at 37 °C for 30 min. PCR was subsequently performed as previously described [25]. Libraries were purified with AMPure beads (Agencourt) to remove contaminating primer dimers. All libraries were sequenced with 50 bp paired-end reads on an Illumina HiSeq 2500 platform. For the control (*A. niger* CBS513.88 naked genome DNA), 5 ng of genomic DNA was used for the Tn5 transposase reaction. All the ATAC-seq data were submitted to the NCBI (Accession number, PRJNA566304, PRJNA587805 and PRJNA692847)

#### Construction of transcription factor knockout and overexpression strains in *Aspergillus*

Different TFs, *creA* [47], *amyR* [65], *cpcA* [69], *pacC* [64], *prtT* [35], and *laeA* [37, 38], were searched on the *Aspergillus* genome database ([www.aspgd.com](http://www.aspgd.com)), and

homologous arms for gene knockout were designed based on the 1000-bp sequence flanking the CDS region of each gene, which were then amplified with primer sets listed in Additional file 12: Table S11. The uridine auxotrophic marker *pyrG* from *A. nidulans* was used as a screening marker for TF knockout. The upstream and downstream homologous arms of each TF obtained by PCR amplification were introduced along with the *pyrG* marker into a pMD20 T-Vector (TaKaRa, Japan) by NEBuilder (New England Biolabs, America), to obtain TF knockout vectors (pdelcreA, pdelamyR, pdelcpcA, pdelpacC, pdelprtT, and pdellaA). The knockout cassettes that were linearized with *XbaI*, *EcoRI*, *ApaI*, *EcoRI*, *EcoRI*, and *BamHI* were introduced into *A. niger* SH2 and *A. oryzae* niaD300 by PEG-mediated protoplast transformation. In brief, protoplasts were prepared as the description in ATAC-seq library preparation. Protoplasts were resuspended in a certain amount of STC, and transformation system was made up of 100 µL linearized plasmid, 160 µL protoplasts, and 60 µL PEG. The mixture was incubated on the ice for 30 min, and then added another 1.5 mL PEG for 25 min at the room temperature. Finally, the mixture was pooled down on the hyperosmotic sucrose CD plate. After 5 days cultivated at 30 °C, transformants were obtained and identified. The cassettes for *laeA* overexpression digested with *ApaI* were transformed into *A. oryzae* niaD300 to yield the OElaeA strain. The TF knockout mutants in which homologous integration occurred were analyzed by PCR amplification for upstream and downstream localization, and they were further validated by qRT-PCR (Additional file 1: Figure S11). The primers used in this study are listed in Additional file 12: Table S11.

#### Chromatin immunoprecipitation and ChIP-seq library preparations

To confirm the ATAC-seq data, we constructed PrtT-FLAG (CBS513.88  $\Delta$ *kusA*; *prtT*-3 × FLAG;*pyrG*) strain, in which the TF PrtT was drove by PglA and labelled by 3 × FLAG tagging, for ChIP-seq analysis. Fungal conidia (10<sup>6</sup>/mL) were incubated in 100 mL DPY media at 30 °C under shaking at 200 rpm for 40 h. The mycelia was harvested, dried with two layers of miracloth, and transferred into buffer A for cross-linking. Cross-linking was done according to the protocol of [12]. Then samples were frozen in liquid nitrogen and lysed with Bullet Blender with the setting 12, 5 × 3 min with a 1-min ice interval in between. The lysed supernatant was collected for sonication (UCD 200 for 30 min) in ice water slurry with the program: 10 S on, 15 S off. Specific antibody anti-FLAG (Sigma, Anti-FLAG M2) was used for binding with the DNA/PrtT complex. For each standard IP, 50 µL chromatin, 2 µg antibody, and 450 µL lysis buffer were added and were kept on a magnetic stand.



Precipitated sample was then washed and the immunoprecipitated DNA was digested with RNase A and purified with DNA extract kit (Qiagen, Cat. No. 27104). The extracted DNA was prepared using Illumina multiplex system and sequenced by Illumina HiSeq 2500 (Life Technologies).

#### RNA extraction, qRT-PCR, and RNA-seq analysis

A small portion of the mycelium used to prepare the ATAC-seq library was left for RNA extraction. Total RNA was isolated using a HiPure Fungal RNA Kit (Magen, China). Reverse transcription was performed with a PrimeScript RT-PCR Kit (TaKaRa, Japan). Quantitative real-time PCR (qRT-PCR) was performed using an ABI 7500 Fast Real-Time PCR System to analyze TFs and the report gene *goxC*; the primers used are listed in Additional file 12: Table S11. The resulting libraries were sequenced at the Beijing Genomics Institute (BGI) with 50 bp single-end reads on the BGISEQ-500 (NCBI accession number, PRJNA588127). The RNA quality was assessed with an Agilent Bioanalyzer 2100 system to confirm that all samples had an RNA integrity value (above 7.0). All samples were prepared in duplicate. To obtain FASTQ format clean reads, all generated raw sequencing reads were filtered to remove reads with adaptors, reads in which unknown bases were greater than 10%, and low-quality reads. After filtering, clean reads were mapped to the genome and cDNA using HISAT [70] and Bowtie2 [71]. The gene expression level was quantified by RSEM [72], and the FPKM value of all samples is provided in Additional file 13: Supplementary Table S12. The DESeq2 method [73] was used to screen differentially expressed genes between the two groups.

#### Mapping and normalization of ATAC-seq data

After removing adaptors using Trimmomatic [74], 50 bp paired-end ATAC-seq reads were mapped to the *Aspergillus* genome (Ensembl/ASM285 for *A. niger* and Ensembl/ASM18445 for *A. oryzae*) using Bowtie2 with parameters “--trim5 5 --trim3 15” and other default parameters. All reads aligned to the + strand and the – strand were offset by  $\pm 5$  bp, respectively, due to the Tn5 transposase introducing two cuts that were separated by 9 bp [75]. Mapped reads of SAM output were converted to a BAM format and sorted by Samtools [76]. Duplicate reads were removed using the default parameters of the Picard tools MarkDuplicates program (<http://broadinstitute.github.io/picard/>).

To visualize mapped reads, BAM format files were converted to bigwig format using the *bamCoverage* tool in deepTools2 with a bin size of 1 bp and with RPKM normalization. Heatmaps and average plots from the ATAC-seq data matrix were also generated using the “*computeMatrix*,” “*plotHeatmap*,” and “*plotProfile*”

functions in the deepTools2 package [77]. Genome browser images were made using the integrative Genomics Viewer (IGV) [78] and bigwig files that were processed as described above. Fragments length between the pair-end adaptors were calculated with the program: `Samtools view ATAC_sorted.bam | awk '$9>0' | cut -f 9 | sort | uniq -c | sort -b -k2, 2n | sed -e 's/^[ \t]*//'` > `fragment_length_count.txt` and the desired length was extracted from SAM files by Python.

#### Peak calling identified accessible chromatin regions

Accessible regions and peaks of each sample were identified using MACS2 [79] with the following parameters: `-f BAMPE --nomodel --shift 100 --extsize 200`. The narrowPeak files from the MACS2 output were used for further analysis. The *annotatePeaks.pl* tool of the HOMER program [45] was used with the default parameters to annotate the location of the identified peaks overlapping with genomic features, the TSS (transcription start site), the TTS (transcription termination site), exons, introns, and intergenic regions, all of which were annotated in the Ensembl *Aspergillus niger* CBS513.88 genome. Peaks that were 1 kb upstream of the TSS were associated with the nearest genes. These genes were then analyzed for overrepresented gene ontology and KEGG pathway using FungiFun 2.2.8 BETA (<https://sbi.hki-jena.de/fungifun/>) [80].

#### Analysis of differential chromatin accessibility

To identify differentially mapped reads, we used Diffbind [81]. We used the processed ATAC-seq alignment bam file from Bowtie2 and the narrowPeak file from MACS2 for each sample. MA plots (log<sub>2</sub>-fold change vs. mean average) were used to visualize changes in chromatin accessibility for all peaks.

#### Identification of footprints

Footprints were identified with pyDNase [44] using the default parameters. The Wellington footprinting algorithms [44] in pyDNase searched for TF footprints in areas of the genome that MACS2 identified as peaks. Wellington applied a beta-binomial distribution to estimate footprints and elevated the strand-specific activity of DNase or Tn5 transposase around transcription factor footprints among accessible chromatin regions.

#### Design and synthesis of a minimal *Aspergillus* promoter for functional verification of CREs in vivo

With reference to the yeast minimal promoter screening system [62], we designed a minimal *Aspergillus* promoter for functional verification of CREs in vivo (Additional file 1: Figure S9a). The 67 bp 5' UTR of the *Aspergillus nidulans* *TPI* gene [82] was used as a TSS. The TATA box and the TSS are separated by a 29-bp distance from

the TSS in the *amyB* (An05g02100) promoter of *A. niger*, which facilitates successful loading of the transcription initiation complex, thereby initiating transcription by RNA polymerase. To ensure that the TATA box can be successfully opened during the transcription process, we manually added 8 AT-enriched sequences in front of the TA box to reduce the bond energy of the DNA [62]. These elements were linked together by fusion PCR amplification to form a core promoter region (Additional file 1: Figure S9a). The glucose oxidase *goxC* (An01g14740) was used as the reporter gene, which could catalyze the oxidation of D-glucose at carbon 1 into D-glucono-1, 5-lactone and hydrogen peroxide. Then the peroxidase catalyzes the decomposition of hydrogen peroxide while oxidizing O-anisidine, thus resulting in a red phenotype [83] (Additional file 1: Figure S9b). To further refine the entire reporting system, the selection marker *pyrG* was integrated before the footprints to separate the footprints from other genomic sequences, ensuring that transcription of the core sequence was not affected by other sequences. The footprints contained a transcription factor binding site (TFBS) that helped stabilize RNAP to enhance transcription efficiency. By constantly changing footprints and monitoring the transcription of downstream reporter genes, it was possible to identify whether the footprints containing the TFBS function in vivo. The 1 kb flanking sequence of the glucoamylase gene (*glaA*) acts as a homologous arm, and the aim was that the sequence successfully integrated into the glucoamylase gene locus, ensuring that all open reading frames were at the same position in the genome [84]. All the report plasmids were linearized with *XbaI* and transformed into *A. niger* by PEG-mediated protoplast transformation.

#### Discovery of de novo transcription factor motifs

ATAC-seq footprints that were identified from each sample were used to discover de novo transcription factor motifs. The boundaries of the start and stop coordinates of ATAC-seq footprints were lengthened by 10 bp upstream and downstream. We then used the *findMotifsGenome.pl* tool of the HOMER program to identify overrepresented transcription factor sites (transcription factor motifs). Using the *annotatePeaks.pl* tool of HOMER with the *-m* parameter, HOMER de novo motifs were then used to search the original set of sequences to identify all instances of motifs.

#### Analysis of known motif enrichment within footprints

The known motif files in the HOMER format were created manually with the *seq2profile.pl* tool of HOMER. The known motifs were then used to search the footprints of each sample to identify all instances of motifs

using the *annotatePeaks.pl* tool of HOMER with the *-m* parameter.

#### Abbreviations

ATAC-seq: Assay for transposase-accessible chromatin using sequencing; TF: Transcription factor; TFBS: Transcription factor binding sites; ChIP-seq: Chromatin immunoprecipitation sequencing; CREs: *cis*-regulatory elements; TSS: Transcription start site; TTS: Transcription terminator site; SMs: Secondary metabolites; PKSs: Polyketide synthases; NRPSs: Non-ribosomal peptide synthetases; AF: Aflatoxin; DEGs: Differential expression genes; OC: Open chromatin; CCR: Carbon catabolite repression

#### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12915-021-01114-0>.

- Additional file 1: Figure S1.** Fragment length distribution of Tn5 transposase integration. **Figure S2.** qRT-PCR analysis of *A. niger* cultured in different condition. **Figure S3.** Enrichment of PrtT binding motif from the ChIP-seq data. **Figure S4.** Visually display of the PrtT binding site. **Figure S5.** Clustering analysis of ATAC-seq peaks for *Aspergillus oryzae* niaD300 and its *laeA* mutants. **Figure S6.** Comparison of ATAC-seq peaks and RNA-seq signal among *A.oryzae* niaD300 and *laeA* mutant. **Figure S7.** *De novo* motifs of SH2 TF-deficient strains. **Figure S8.** GO analysis of 3 *de novo* predicted over-represented TFBS targeting genes. **Figure S9.** The minimal promoter structure and the verification of its function. **Figure S10.** *In vivo* functional verification of AreA targeting sites. **Figure S11.** qRT-PCR verification of TF knockout strains of *A.niger* SH2 and *A.oryzae* niaD300.
- Additional file 2: Supplementary Table S1.** Sequencing data quality and mapping rate analysis of the ATAC-seq samples.
- Additional file 3: Supplementary Table S2.** Chromatin accessible regions (Peaks) of each sample identified by MACS2.
- Additional file 4: Supplementary Table S3.** Syntenic orthologues of ATAC-seq peaks identified between *A. niger* CB513.88 and *A. oryzae* niaD300.
- Additional file 5: Supplementary Table S4.** Differential peaks analysis of *A. niger* and its mutants.
- Additional file 6: Supplementary Table S5.** Differential peaks and SM cluster analysis of *A. oryzae* and its *laeA* mutants.
- Additional file 7: Supplementary Table S6.** DNA footprints of *A. niger* wild-type strains and TF-deleted strains.
- Additional file 8: Supplementary Table S7.** *De novo* TFBS of *A. niger* wild-type strains and TF-deleted strains with a *P* value threshold of 1E-50 by HOMER.
- Additional file 9: Supplementary Table S8.** The physical position and gene function of DNA instances containing the TFBS mapped to the *A. niger* genome.
- Additional file 10: Supplementary Table S9.** Motif instances of known TFs from footprints by comparing *A. niger* WT strains and the TF-deleted strains.
- Additional file 11: Supplementary Table S10.** Genotype of strains used in this study.
- Additional file 12: Supplementary Table S11.** Primers used in this study
- Additional file 13: Supplementary Table S12.** FPKM value of all the *A. niger* RNA-seq samples.
- Additional file 14: Supplementary Table S13.** Original data of all the figures and additional files.

#### Acknowledgements

We are very grateful to Vazyme Biotech Co., Ltd (Nanjing, China) for providing the Tn5 reaction kit and the corresponding technicians. We also thank BGI Company for NGS services.

### Authors' contributions

LH was responsible for conception and design, acquisition of the data, analysis and interpretation of the data, and drafting of the article. XL, LY, and LD were responsible for acquisition of the data, analysis, and interpretation of the data. BW and LP responsible for conception and design, acquisition of the data, analysis and interpretation of the data, and drafting of the article. All authors read and approved the final manuscript.

### Funding

This work was supported by the National Natural Science Foundation of China (grant number 31871736, 31870024), the Science and Technology Planning Project of Tianjin City (grant no. TSBICIP-KJGG-006-21), the Core Technology Project of Foshan City (project no.1920001000824), and the Science and Technology Planning Project of Guangzhou City (grant number 202002030207).

### Availability of data and materials

All data generated or analyzed during this study are included in this published article, its supplementary information files, and publicly available repositories. The raw ATAC-seq and ChIP-seq sequencing data have been deposited in the NCBI SRA database under the accession number PRJNA566304, PRJNA587805, PRJNA692847, and PRJNA692789. The raw RNA-seq sequencing data have been deposited in the NCBI SRA database under the accession number PRJNA588127. The Supporting data are included in the additional files 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 and 13. A file providing individual data values was submitted as Additional file 14.

### Declarations

#### Ethics approval and consent to participate

Not applicable.

#### Consent for publication

Not applicable.

#### Competing interests

The authors declare that no competing interests exist.

### Author details

<sup>1</sup>School of Biology and Biological Engineering, South China University of Technology, Guangzhou Higher Education Mega Center, Guangzhou 510006, China. <sup>2</sup>Guangdong Provincial Key Laboratory of Fermentation and Enzyme Engineering, South China University of Technology, Guangzhou Higher Education Mega Center, Guangzhou 510006, China.

Received: 8 January 2020 Accepted: 2 August 2021

Published online: 06 September 2021

### References

- Andersen MR, Salazar MP, Schaap PJ, van de Vondervoort PJ, Culley D, Thykaer J, et al. Comparative genomics of citric-acid-producing *Aspergillus niger* ATCC 1015 versus enzyme-producing CBS 513.88. *Genome Res.* 2011; 21(6):885–97. <https://doi.org/10.1101/gr.112169.110>.
- Vesth TC, Nybo JL, Theobald S, Frisvad JC, Larsen TO, Nielsen KF, et al. Investigation of inter- and intraspecies variation through genome sequencing of *Aspergillus* section Nigri. *Nat Genet.* 2018;50(12):1688–95. <https://doi.org/10.1038/s41588-018-0246-1>.
- Pel HJ, de Winde JH, Archer DB, Dyer PS, Hofmann G, Schaap PJ, et al. Genome sequencing and analysis of the versatile cell factory *Aspergillus niger* CBS 513.88. *Nat Biotechnol.* 2007;25(2):221–31. <https://doi.org/10.1038/nbt1282>.
- Yin C, Wang B, He P, Lin Y, Pan L. Genomic analysis of the aconidial and high-performance protein producer, industrially relevant *Aspergillus niger* SH2 strain. *Gene.* 2014;541(2):107–14. <https://doi.org/10.1016/j.gene.2014.03.011>.
- Todd RB, Zhou M, Ohm RA, Leeggangers HA, Visser L, de Vries RP. Prevalence of transcription factors in ascomycete and basidiomycete fungi. *BMC Genomics.* 2014;15(1):214. <https://doi.org/10.1186/1471-2164-15-214>.
- Schape P, Kwon MJ, Baumann B, Gutschmann B, Jung S, Lenz S, et al. Updating genome annotation for the microbial cell factory *Aspergillus niger* using gene co-expression networks. *Nucleic Acids Res.* 2019;47(2):559–69. <https://doi.org/10.1093/nar/gky1183>.
- Felsenfeld G. Chromatin as an essential part of the transcriptional mechanism. *Nature.* 1992;355(6357):219–24. <https://doi.org/10.1038/355219a0>.
- Qiu Z, Li R, Zhang S, Wang K, Xu M, Li J, et al. Identification of regulatory DNA elements using genome-wide mapping of DNase I hypersensitive sites during tomato fruit development. *Mol Plant.* 2016;9(8):1168–82. <https://doi.org/10.1016/j.molp.2016.05.013>.
- Galagan JE, Calvo SE, Cuomo C, Ma LJ, Wortman JR, Batzoglou S, et al. Sequencing of *Aspergillus nidulans* and comparative analysis with *A. fumigatus* and *A. oryzae*. *Nature.* 2005;438(7071):1105–15. <https://doi.org/10.1038/nature04341>.
- Johnson DS, Mortazavi A, Myers RM, Wold B. Genome-wide mapping of *in vivo* protein-DNA interactions. *Science.* 2007;316(5830):1497–502. <https://doi.org/10.1126/science.1141319>.
- Park PJ. ChIP-seq: advantages and challenges of a maturing technology. *Nat Rev Genet.* 2009;10(10):669–80. <https://doi.org/10.1038/nrg2641>.
- de Castro PA, Chen C, de Almeida RS, Freitas FZ, Bertolini MC, Morais ER, et al. ChIP-seq reveals a role for CrzA in the *Aspergillus fumigatus* high-osmolarity glycerol response (HOG) signalling pathway. *Mol Microbiol.* 2014; 94(3):655–74. <https://doi.org/10.1111/mmi.12785>.
- Chung D, Barker BM, Carey CC, Merriman B, Werner ER, Lechner BE, et al. ChIP-seq and *in vivo* transcriptome analyses of the *Aspergillus fumigatus* SREBP SrbA reveals a new regulator of the fungal hypoxia response and virulence. *PLoS Pathog.* 2014;10(11):e1004487. <https://doi.org/10.1371/journal.ppat.1004487>.
- Amich J, Schaffner L, Haas H, Krappmann S. Regulation of sulphur assimilation is essential for virulence and affects iron homeostasis of the human-pathogenic mould *Aspergillus fumigatus*. *PLoS Pathog.* 2013;9(8): e1003573. <https://doi.org/10.1371/journal.ppat.1003573>.
- Paul S, Starnes M, Thomas GH, Liu H, Hagiwara D, Gomi K, et al. AtRr is an essential determinant of azole resistance in *Aspergillus fumigatus*. *mBio.* 2019;10(2). <https://doi.org/10.1128/mbio.02563-18>.
- Wu MY, Mead ME, Lee MK, Ostrem Loss EM, Kim SC, Rokas A, et al. Systematic dissection of the evolutionarily conserved WetA developmental regulator across a genus of filamentous fungi. *mBio.* 2018;9(4). <https://doi.org/10.1128/mBio.01130-18>.
- Sun J, Glass NL. Identification of the CRE-1 cellulolytic regulon in *Neurospora crassa*. *PLoS One.* 2011;6(9):e25654. <https://doi.org/10.1371/journal.pone.0025654>.
- Nasmith CG, Walkowiak S, Wang L, Leung WW, Gong Y, Johnston A, et al. Tri6 is a global transcription regulator in the phytopathogen *Fusarium graminearum*. *PLoS Pathog.* 2011;7(9):e1002266. <https://doi.org/10.1371/journal.ppat.1002266>.
- Kim S, Hu J, Oh Y, Park J, Choi J, Lee YH, et al. Combining ChIP-chip and expression profiling to model the MoCRZ1 mediated circuit for Ca/calciurein signaling in the rice blast fungus. *PLoS Pathog.* 2010;6(5): e1000909. <https://doi.org/10.1371/journal.ppat.1000909>.
- Gacek-Matthews A, Berger H, Sasaki T, Wittstein K, Gruber C, Lewis ZA, et al. KdmB, a jumoni histone H3 demethylase, regulates genome-wide H3K4 trimethylation and is required for normal induction of secondary metabolism in *Aspergillus nidulans*. *PLoS Genet.* 2016;12(8):e1006222. <https://doi.org/10.1371/journal.pgen.1006222>.
- Fischer J, Muller SY, Netzker T, Jager N, Gacek-Matthews A, Scherlach K, et al. Chromatin mapping identifies BasR, a key regulator of bacteria-triggered production of fungal secondary metabolites. *eLife.* 2018;7:e40969. <https://doi.org/10.7554/eLife.40969>.
- Giresi PG, Kim J, McDaniell RM, Iyer VR, Lieb JD. FAIRE (formaldehyde-assisted isolation of regulatory elements) isolates active regulatory elements from human chromatin. *Genome Res.* 2007;17(6):877–85. <https://doi.org/10.1101/gr.5533506>.
- Wang C, Lv Y, Wang B, Yin C, Lin Y, Pan L. Survey of protein-DNA interactions in *Aspergillus oryzae* on a genomic scale. *Nucleic Acids Res.* 2015;43(9):4429–46. <https://doi.org/10.1093/nar/gkv334>.
- Zentner GE, Henikoff S. High-resolution digital profiling of the epigenome. *Nat Rev Genet.* 2014;15(12):814–27. <https://doi.org/10.1038/nrg3798>.
- Buenostro JD, Wu B, Chang HY, Greenleaf WJ. ATAC-seq: A method for assaying chromatin accessibility genome-wide. *Curr Protocols Mol Biol.* 2015;109:21 9 1–9.
- Li Z, Schulz MH, Look T, Begemann M, Zenke M, Costa IG. Identification of transcription factor binding sites using ATAC-seq. *Genome Biol.* 2019;20(1): 45. <https://doi.org/10.1186/s13059-019-1642-2>.

27. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods*. 2013;10(12):1213–8. <https://doi.org/10.1038/nmeth.2688>.
28. Wu J, Huang B, Chen H, Yin Q, Liu Y, Xiang Y, et al. The landscape of accessible chromatin in mammalian preimplantation embryos. *Nature*. 2016; 534(7609):652–7. <https://doi.org/10.1038/nature18606>.
29. Fernandez-Minan A, Bessa J, Tena JJ, Gomez-Skarmeta JL. Assay for transposase-accessible chromatin and circularized chromosome conformation capture, two methods to explore the regulatory landscapes of genes in zebrafish. *Methods Cell Biol*. 2016;135:413–30. <https://doi.org/10.1016/bs.mcb.2016.02.008>.
30. Bozek M, Cortini R, Storti AE, Unnerstall U, Gaul U, Gompel N. ATAC-seq reveals regional differences in enhancer accessibility during the establishment of spatial coordinates in the *Drosophila blastoderm*. *Genome Res*. 2019;29(5):771–83. <https://doi.org/10.1101/gr.242362.118>.
31. Daugherty AC, Yeo RW, Buenrostro JD, Greenleaf WJ, Kundaje A, Brunet A. Chromatin accessibility dynamics reveal novel functional enhancers in *C. elegans*. *Genome Res*. 2017;27(12):2096–107. <https://doi.org/10.1101/gr.226233.117>.
32. Tannenbaum M, Sarusi-Portuguez A, Krispil R, Schwartz M, Loza O, Benichou JIC, et al. Regulatory chromatin landscape in *Arabidopsis thaliana* roots uncovered by coupling INTACT and ATAC-seq. *Plant Methods*. 2018;14(1): 113. <https://doi.org/10.1186/s13007-018-0381-9>.
33. Schep AN, Buenrostro JD, Denny SK, Schwartz K, Sherlock G, Greenleaf WJ. Structured nucleosome fingerprints enable high-resolution mapping of chromatin architecture within regulatory regions. *Genome Res*. 2015;25(11): 1757–70. <https://doi.org/10.1101/gr.192294.115>.
34. Mavrich TN, Ioshikhes IP, Venters BJ, Jiang C, Tomsho LP, Qi J, et al. A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome. *Genome Res*. 2008;18(7):1073–83. <https://doi.org/10.1101/gr.078261.108>.
35. Caruso ML, Litzka O, Martic G, Lottspeich F, Brakhage AA. Novel basic-region helix-loop-helix transcription factor (AnBH1) of *Aspergillus nidulans* counteracts the CCAAT-binding complex AnCF in the promoter of a penicillin biosynthesis gene. *J Mol Biol*. 2002;323(3):425–39. [https://doi.org/10.1016/S0022-2836\(02\)00965-8](https://doi.org/10.1016/S0022-2836(02)00965-8).
36. Bok JW, Keller NP. LaeA, a regulator of secondary metabolism in *Aspergillus* spp. *Eukaryot Cell*. 2004;3(2):527–35. <https://doi.org/10.1128/EC.3.2.527-535.2004>.
37. Bok JW, Hoffmeister D, Maggio-Hall LA, Murillo R, Glasner JD, Keller NP. Genomic mining for *Aspergillus* natural products. *Chem Biol*. 2006;13(1):31–7. <https://doi.org/10.1016/j.chembiol.2005.10.008>.
38. Wang B, Lv Y, Li X, Lin Y, Deng H, Pan L. Profiling of secondary metabolite gene clusters regulated by LaeA in *Aspergillus niger* FGSC A1279 based on genome sequencing and transcriptome analysis. *Res Microbiol*. 2018;169(2): 67–77. <https://doi.org/10.1016/j.resmic.2017.10.002>.
39. Patananan AN, Palmer JM, Garvey GS, Keller NP, Clarke SG. A novel automethylation reaction in the *Aspergillus nidulans* LaeA protein generates S-methylmethionine. *J Biol Chem*. 2013;288(20):14032–45. <https://doi.org/10.1074/jbc.M113.465765>.
40. Khaldi N, Seifuddin FT, Turner G, Haft D, Nierman WC, Wolfe KH, et al. SMURF: Genomic mapping of fungal secondary metabolite clusters. *Fungal Genet Biol*. 2010;47(9):736–41. <https://doi.org/10.1016/j.fgb.2010.06.003>.
41. Karabacak Calviello A, Hirsekorn A, Wurms R, Yusuf D, Ohler U. Reproducible inference of transcription factor footprints in ATAC-seq and DNase-seq datasets using protocol-specific bias modeling. *Genome Biol*. 2019;20(1):42. <https://doi.org/10.1186/s13059-019-1654-y>.
42. Lu Z, Hofmeister BT, Vollmers C, DuBois RM, Schmitz RJ. Combining ATAC-seq with nuclei sorting for discovery of cis-regulatory regions in plant genomes. *Nucleic Acids Res*. 2017;45(6):e41. <https://doi.org/10.1093/nar/gkw1179>.
43. Piper J, Elze MC, Cauchy P, Cockerill PN, Bonifer C, Ott S. Wellington: a novel method for the accurate identification of digital genomic footprints from DNase-seq data. *Nucleic Acids Res*. 2013;41(21):e201. <https://doi.org/10.1093/nar/gkt850>.
44. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell*. 2010;38(4):576–89. <https://doi.org/10.1016/j.molcel.2010.05.004>.
45. Zhou B, Wang C, Wang B, Li X, Xiao J, Pan L. Identification of functional cis-elements required for repression of the Taka-amylase A gene under secretion stress in *Aspergillus oryzae*. *Biotechnol Lett*. 2015;37(2):333–41. <https://doi.org/10.1007/s10529-014-1691-2>.
46. Drysdale MR, Kolze SE, Kelly JM. The *Aspergillus niger* carbon catabolite repressor encoding gene, *creA*. *Gene*. 1993;130(2):241–5. [https://doi.org/10.1016/0378-1119\(93\)90425-3](https://doi.org/10.1016/0378-1119(93)90425-3).
47. Ackermann AM, Wang Z, Schug J, Naji A, Kaestner KH. Integration of ATAC-seq and RNA-seq identifies human alpha cell and beta cell signature genes. *Mol Metab*. 2016;5(3):233–44. <https://doi.org/10.1016/j.molmet.2016.01.002>.
48. Toenhake CG, Frascchka SA, Vijayabaskar MS, Westhead DR, van Heeringen SJ, Bartfai R. Chromatin accessibility-based characterization of the gene regulatory network underlying plasmodium falci-parum blood-stage development. *Cell Host Microbe*. 2018;23(4):557–69 e9. <https://doi.org/10.1016/j.chom.2018.03.007>.
49. Chen L, Zou G, Zhang L, de Vries RP, Yan X, Zhang J, et al. The distinctive regulatory roles of PrtT in the cell metabolism of *Penicillium oxalicum*. *Fungal Genet Biol*. 2014;63:42–54. <https://doi.org/10.1016/j.fgb.2013.12.001>.
50. Huang L, Dong L, Wang B, Pan L. The transcription factor PrtT and its target protease profiles in *Aspergillus niger* are negatively regulated by carbon sources. *Biotechnol Lett*. 2020;42(4):613–24. <https://doi.org/10.1007/s10529-020-02806-3>.
51. Gusmao EG, Allhoff M, Zenke M, Costa IG. Analysis of computational footprinting methods for DNase sequencing experiments. *Nat Methods*. 2016;13(4):303–9. <https://doi.org/10.1038/nmeth.3772>.
52. Strittmatter AW, Irniger S, Braus GH. Induction of *jbA* mRNA synthesis for a putative bZIP protein of *Aspergillus nidulans* by amino acid starvation. *Curr Genet*. 2001;39(5-6):327–34. <https://doi.org/10.1007/s002940100215>.
53. Balazs A, Pocsí I, Hamari Z, Leiter E, Emri T, Miskei M, et al. AtfA bZIP-type transcription factor regulates oxidative and osmotic stress responses in *Aspergillus nidulans*. *Mol Gen Genomics*. 2010;283(3):289–303. <https://doi.org/10.1007/s00438-010-0513-z>.
54. Mendoza-Martinez AE, Lara-Rojas F, Sanchez O, Aguirre J. NapA mediates a redox regulation of the antioxidant response, carbon utilization and development in *Aspergillus nidulans*. *Front Microbiol*. 2017;8:516. <https://doi.org/10.3389/fmicb.2017.00516>.
55. Sakamoto K, Arima TH, Iwashita K, Yamada O, Gomi K, Akita O. *Aspergillus oryzae atfB* encodes a transcription factor required for stress tolerance in conidia. *Fungal Genet Biol*. 2008;45(6):922–32. <https://doi.org/10.1016/j.fgb.2008.03.009>.
56. Sakamoto K, Iwashita K, Yamada O, Kobayashi K, Mizuno A, Akita O, et al. *Aspergillus oryzae atfA* controls conidial germination and stress tolerance. *Fungal Genet Biol*. 2009;46(12):887–97. <https://doi.org/10.1016/j.fgb.2009.09.004>.
57. Tian C, Li J, Glass NL. Exploring the bZIP transcription factor regulatory network in *Neurospora crassa*. *Microbiology*. 2011;157(Pt 3):747–59. <https://doi.org/10.1099/mic.0.045468-0>.
58. Jin FJ, Takahashi T, Matsushima K, Hara S, Shinohara Y, Maruyama J, et al. SclR, a basic helix-loop-helix transcription factor, regulates hyphal morphology and promotes sclerotial formation in *Aspergillus oryzae*. *Eukaryot Cell*. 2011;10(7):945–55. <https://doi.org/10.1128/EC.00013-11>.
59. Fisher F, Goding CR. Single amino acid substitutions alter helix-loop-helix protein specificity for bases flanking the core CANNTG motif. *EMBO J*. 1992; 11(11):4103–9. <https://doi.org/10.1002/j.1460-2075.1992.tb05503.x>.
60. Zhou X, O'Shea EK. Integrated approaches reveal determinants of genome-wide binding and function of the transcription factor Pho4. *Mol Cell*. 2011; 42(6):826–36. <https://doi.org/10.1016/j.molcel.2011.05.025>.
61. Redden H, Alper HS. The development and characterization of synthetic minimal yeast promoters. *Nat Commun*. 2015;6(1):7810. <https://doi.org/10.1038/ncomms8810>.
62. Carvalho ND, Jorgensen TR, Arentshorst M, Nitsche BM, van den Hondel CA, Archer DB, et al. Genome-wide expression analysis upon constitutive activation of the HacA bZIP transcription factor in *Aspergillus niger* reveals a coordinated cellular response to counteract ER stress. *BMC Genomics*. 2012; 13(1):350. <https://doi.org/10.1186/1471-2164-13-350>.
63. MacCabe AP, Van den Hombergh JP, Tilburn J, Arst HN Jr, Visser J. Identification, cloning and analysis of the *Aspergillus niger* gene *pacC*, a wide domain regulatory gene responsive to ambient pH. *Mol Gen Genet*. 1996;250(3):367–74. <https://doi.org/10.1007/BF02174395>.
64. Punt PJ, Schuren FH, Lehmbeck J, Christensen T, Hjort C, van den Hondel CA. Characterization of the *Aspergillus niger prtT*, a unique regulator of

- extracellular protease encoding genes. *Fungal Genet Biol.* 2008;45(12):1591–9. <https://doi.org/10.1016/j.fgb.2008.09.007>.
65. vanKuyk PA, Benen JA, Wosten HA, Visser J, de Vries RP. A broader role for AmyR in *Aspergillus niger*: regulation of the utilisation of D-glucose or D-galactose containing oligo- and polysaccharides. *Appl Microbiol Biotechnol.* 2012;93(1):285–93. <https://doi.org/10.1007/s00253-011-3550-6>.
  66. Suzuki K, Tanaka M, Konno Y, Ichikawa T, Ichinose S, Hasegawa-Shiro S, et al. Distinct mechanism of activation of two transcription factors, AmyR and MalR, involved in amyolytic enzyme production in *Aspergillus oryzae*. *Appl Microbiol Biotechnol.* 2015;99(4):1805–15. <https://doi.org/10.1007/s00253-014-6264-8>.
  67. Strauss J, Horvath HK, Abdallah BM, Kindermann J, Mach RL, Kubicek CP. The function of CreA, the carbon catabolite repressor of *Aspergillus nidulans*, is regulated at the transcriptional and post-transcriptional level. *Mol Microbiol.* 1999;32(1):169–78. <https://doi.org/10.1046/j.1365-2958.1999.01341.x>.
  68. Dong L, Yu D, Lin X, Wang B, Pan L. Improving expression of thermostable trehalase from *Myceliophthora sepedonium* in *Aspergillus niger* mediated by the CRISPR/Cas9 tool and its purification, characterization. *Protein Expr Purif.* 2019;165:105482.
  69. Wanke C, Eckert S, Albrecht G, van Hartingsveldt W, Punt PJ, van den Hondel CAMJJ, et al. The *Aspergillus niger* GCN4 homologue, *cpcA*, is transcriptionally regulated and encodes an unusual leucine zipper. *Mol Microbiol.* 1997;23(1):23–33. <https://doi.org/10.1046/j.1365-2958.1997.1741549.x>.
  70. Kim D, Landmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods.* 2015;12(4):357–U121. <https://doi.org/10.1038/nmeth.3317>.
  71. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* 2012;9(4):357–9. <https://doi.org/10.1038/nmeth.1923>.
  72. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics.* 2011;12(1):323. <https://doi.org/10.1186/1471-2105-12-323>.
  73. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15(12):550. <https://doi.org/10.1186/s13059-014-0550-8>.
  74. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 2014;30(15):2114–20. <https://doi.org/10.1093/bioinformatics/btu170>.
  75. Adey A, Morrison HG, Asan, Xun X, Kitzman JO, Turner EH, et al. Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition. *Genome Biol.* 2010;11(12):R119. <https://doi.org/10.1186/gb-2010-11-12-r119>.
  76. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. *Bioinformatics.* 2009;25(16):2078–9. <https://doi.org/10.1093/bioinformatics/btp352>.
  77. Ramirez F, Ryan DP, Gruning B, Bhardwaj V, Kilpert F, Richter AS, et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* 2016;44(W1):W160–W5. <https://doi.org/10.1093/nar/gkw257>.
  78. Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform.* 2013;14(2):178–92. <https://doi.org/10.1093/bib/bbs017>.
  79. Zhang Y, Liu T, Meyer CA, Eickhout J, Johnson DS, Bernstein BE, et al. Model-based Analysis of ChIP-Seq (MACS). *Genome Biol.* 2008;9(9):R137. <https://doi.org/10.1186/gb-2008-9-9-r137>.
  80. Priebe S, Kreisel C, Horn F, Guthke R, Linde J. FungiFun2: a comprehensive online resource for systematic analysis of gene lists from fungal species. *Bioinformatics.* 2015;31(3):445–6. <https://doi.org/10.1093/bioinformatics/btu627>.
  81. Ross-Innes CS, Stark R, Teschendorff AE, Holmes KA, Ali HR, Dunning MJ, et al. Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature.* 2012;481(7381):389–U177. <https://doi.org/10.1038/nature10730>.
  82. McKnight GL, O'Hara PJ, Parker ML. Nucleotide sequence of the triosephosphate isomerase gene from *Aspergillus nidulans*: implications for a differential loss of introns. *Cell.* 1986;46(1):143–7. [https://doi.org/10.1016/0092-8674\(86\)90868-8](https://doi.org/10.1016/0092-8674(86)90868-8).
  83. Guo Y, Lu F, Zhao H, Tang Y, Lu Z. Cloning and heterologous expression of glucose oxidase gene from *Aspergillus niger* Z-25 in *Pichia pastoris*. *Appl Biochem Biotechnol.* 2010;162(2):498–509. <https://doi.org/10.1007/s12010-009-8778-6>.
  84. Rantasalo A, Landowski CP, Kuivanen J, Korppoo A, Reuter L, Koivistoinen O, et al. A universal gene expression system for fungi. *Nucleic Acids Res.* 2018;46(18):e111. <https://doi.org/10.1093/nar/gky558>.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

