implemented in the k-means function of R v3.3.1 was used on the TPM value for each gene in each sample for all differentially expressed genes to identify clusters of genes with the same transcriptional response. The appropriate number of clusters to use was determined by plotting the WSS (within-cluster sum of squares) as a function of the number of clusters and chosing the number of clusters beyond which the intra-cluster variation no longer decreased. Heat maps of the different k-means clusters were drawn using the $\log_2$(TPM) value for each gene using the gplots v3.0.1 R package. Enriched InterPro descriptions and GO terms within each cluster were defined using a Fisher's exact test *P*-value <0.01 as implemented in the fisher.test function in R v3.3.1. A principal component analysis (PCA) was done with the prcomp function in the R stats package using the TPM values for each of the genes in all samples. Each individual sample was labelled using different colours while time points were differentiated using shapes. Genes contributing to the first and second principal components were extracted using the R package factoextra v1.0.5 and enriched functional terms were determined as described above.

Differential expression analysis between specific experimental groups was performed using the EdgeR package from Bioconductor [25]. For these specific comparisons, a gene was considered differentially expressed if the FDR for differential expression was less than 0.01 and the absolute $\log_2$ fold-change (LFC) was greater than or equal to one. The circular displays of the significant LFC values were generated using Circos 0.69–5 [26].

GO term and Metabolic Pathway enrichment analyses were performed using the tools available at FungiDB (http://fungidb.org/fungidb/) using default settings [27]. VENNY2.1 was used to generate the Venn diagrams (http://bioinfogp.cnb.csic.es/tools/venny/)

RESULTS AND DISCUSSION

RNA-seq of i*n vitro A. fumigatus* infections

To understand the molecular nature of the interaction between fungal pathogen and host cells in the context of aspergillosis, we performed RNA-seq analysis on poly(A)-enriched RNA isolated from monolayers of airway epithelial cells (A549 cells) infected for 6 or 16 h with two well-characterized clinical isolates of *A. fumigatus*, strains Af293 and CEA10.

After being added to human cell culture, the resting conidia first swell and then, after approximately 4 h, begin to germinate. Following germination, the hyphae elongate rapidly and begin to branch such that the host cell monolayer is covered by a hyphal mat by 16 h after addition of the conidia (Fig. S1). This process occurs independently of the presence of host cells, so it is important to differentiate between transcriptional changes in the fungi that are specifically induced by interaction with the host cells from those that are part of a transcription programme that accompanies swelling, germination and mycelial growth. To accomplish this, RNA-seq was performed on time-matched controls in which each isolate was grown in tissue culture medium without host cells. Each of the eight experimental conditions (Table 1) was carried out in triplicate. The RNA preparations contained a mixture of mRNAs expressed by *A. fumigatus* as well as by the host cells. All sequencing reads were aligned to the reference genome (Af293 or CEA10). Analysis of the host transcriptome will be described in a separate publication. From each of the 24 sequencing libraries we obtained an average of 53±29.3 million reads that mapped to the appropriate fungal reference genome (Table S1).

Common and strain-specific responses to the *in vitro* growth conditions

While the genomes of these two *A. fumigatus* strains are highly similar and syntenic, they are not identical; each strain has many genes with no detectable orthologue in the other strain [28, 29]. We took a conservative approach to comparing the transcriptomes by focusing our analyses on the core set of genes that contain clear 1 : 1 homologues in both genomes. This allowed us to avoid strain-specific responses that stem from large-scale differences in the genomes. Using Jaccard cluster analysis, we defined a set of 9041 genes that we refer to as the 'core' genome (see Materials and Methods) (Table S2). In total, 7888 of these core genes were differentially expressed (FDR<0.01, regardless of fold-change size) in at least one experimental group compared to any of the other seven groups. For this set of 7888 genes, we performed hierarchical clustering of all 24 samples using the $\log_2$-transformed TPM values. Clustering analysis demonstrated that the samples grouped primarily by time spent in the tissue culture media, with additional grouping of samples based on strain (Fig. 1a). PCA confirmed the groupings determined by hierarchical clustering (Fig. 1b).

The top genes, contributing to 95 % of the principal components, were subjected to a functional term enrichment analysis. The 163 genes with the highest contribution to the first principal component are overrepresented in functional terms relating to the ribosome and translation (GO:0005840, GO:0006412) (Table S3). Similarly, the 171 genes with the highest contribution to the second principal component were enriched for ribosomal and translational functional terms, albeit less than the first principal

Table 1. Experimental groups

Group	Host cells	*A. fumigatus* strain	Incubation time (h)
1	–	Af293	6
2	A549	Af293	6
3	–	Af293	16
4	A549	Af293	16
5	–	CEA10	6
6	A549	CEA10	6
7	–	CEA10	16
8	A549	CEA10	16