

## **Role of Genomics and RNA-seq in Studies of Fungal Virulence**

### **Keywords:**

*Aspergillus*, bioinformatics, *Candida*, ChIP-seq, *Cryptococcus*, dermatophytes, genomics, next-generation sequencing, RNA-seq.

### **Authors:**

Alessandro Riccombeni and Geraldine Butler (corresponding author)

School of Biomolecular and Biomedical Science, Conway Institute, University

College Dublin, Belfield, Dublin 4, Ireland.

[alessandro.riccombeni@ucd.ie](mailto:alessandro.riccombeni@ucd.ie)

[geraldine.butler@ucd.ie](mailto:geraldine.butler@ucd.ie)

**Abstract:**

Since its introduction in the last decade, massive parallel sequencing, or “next-generation sequencing”, has revolutionized our access to genomic information, providing accurate data with increasingly higher yields and lower costs with respect to first-generation technology. Massive parallel sequencing of cDNA, or RNA-seq, is progressively replacing array-based technology as the method of choice for transcriptomics. This review describes some of the most recent applications of next-generation sequencing technology to the study of pathogenic fungi, including *Candida*, *Aspergillus* and *Cryptococcus* species. Several integrated approaches illustrate the power and accuracy of RNA-seq for studying the biology of human fungal pathogens. In addition, the lack of consistency in data analysis is discussed.

**Introduction**

The development of next generation sequencing technologies has probably had more impact on our understanding of fungal biology than any other eukaryotic species group. By 2012, the genomes of approximately 200 fungal species have been sequenced ([www.fungalgenomes.org](http://www.fungalgenomes.org)).

More recently, RNA-seq, or high-throughput cDNA sequencing technology, has been increasingly adopted as a method of obtaining large amounts of transcriptomic data with high sensitivity, high quality and at relatively low cost [1, 2]. The major advantages over the previously used microarrays is that no prior knowledge of the transcriptome is required; the high sensitivity allows a deeper characterization of the transcriptional landscape, including non coding transcripts, rare and novel transcripts, fusion products and rare splicing isoforms [1]; and strand-specific methods can be used to improve annotation and to identify antisense transcription [3]. The

development of the technologies has been accompanied by significant advances in data analysis [4-7••].

In this review we describe the recent applications of genome sequencing and RNA sequencing to addressing important biological questions in major fungal pathogens.

## **Genomics of pathogenic fungi**

The yeast *Saccharomyces cerevisiae* was the first eukaryotic genome to be fully sequenced, in 1996 [8]. Other notable landmarks followed, including the genomes of *Candida albicans* [9] and 6 other *Candida* species [10], of filamentous fungi such as the *Aspergilli* [11, 12] and of basidiomycete species including *Cryptococcus* [13]. As new genome sequences became available, the application of comparative analysis allowed the identification of gene families with roles in fungal virulence [10] and the identification of potential new lead compounds for drug discovery [14].

*Candida* species are among the most common fungi associated with hospital-acquired infection, particularly of immunocompromised individuals [15]. Whereas most infections are caused by *C. albicans*, other species such as *Candida parapsilosis*, *Candida tropicalis* and the distantly related *Candida glabrata* are increasing causes of concern [16•]. Comparison of the genomes of pathogenic *Candida* species with closely related non-pathogens showed that several gene families associated with virulence are expanded in the pathogens [10]. These include the *ALS* family of adhesins the Hyr/Iff family of cell wall proteins, and other less well-classified families, many associated with the cell wall.

Comparisons of small numbers of closely related species can also contribute to our understanding of virulence. For example, comparing the genome of *C. albicans*

with that of the substantially less pathogenic species *Candida dubliniensis* revealed that there has been substantial gene loss and pseudogenization (loss of gene function through mutation) in the latter species [17]. More recently, Riccombeni et al. [18•] used next generation sequencing to carry out a comparative analysis of *Candida parapsilosis* and *Candida orthopsilosis*.

*C. parapsilosis* is the second most frequent cause of candidiasis in Latin America and is particularly common in Europe [19]. *C. orthopsilosis* is very closely related to *C. parapsilosis*, and was previously identified as a subgroup [20]. Despite their phenotypic similarity, *C. orthopsilosis* is a much less frequent cause of infection, and is responsible for between 1.4% [21] and 23.8% [22] of infections previously attributed to *C. parapsilosis*. Riccombeni et al. [18] used a combination of 454, Illumina and Sanger technologies to sequence the *C. orthopsilosis* genome. The 8 chromosomal scaffolds encode 5700 genes, of which 97.7% have a direct ortholog in *C. parapsilosis*. The divergence time between *C. parapsilosis* and *C. orthopsilosis* is approximately twice that between *C. albicans* and *C. dubliniensis*. Comparison of the four species revealed that the *WOR2* and *CTA26/TLO2* regulators are evolving faster in *C. albicans/C. dubliniensis* than in *C. parapsilosis/C. orthopsilosis*. *WOR2* is a global regulator of the white/opaque switch in *C. albicans*, a morphological change associated with virulence [23], whereas *TLO2* is a component of the mediator complex and is associated with the regulation of filamentation [24]. The rapid evolution of these two genes may be associated with the increased virulence of *C. albicans*.

Interestingly, there has been an expansion of the H<sup>+</sup> Antiporter-1 (*DHA1*) and *FLU1* clades of the Major Facilitator (MFS) family of drug transporters in *C. parapsilosis* and *C. orthopsilosis* with respect to other *Candida* species. This may be

related to the ability of *C. parapsilosis* to rapidly acquire resistance to azole drugs [25]. In contrast, the Hyr/Iff family is substantially reduced in the less pathogenic *C. orthopsilosis* genome [18••].

The dermatophyte fungi are major causes of superficial infections in human and animal hosts, where they grow within keratinized structures. Common infections include athlete's foot, ringworm and onychomycosis, and infection is particularly high in hot and humid countries [26•]. These species are particularly difficult to study because they grow very slowly under laboratory conditions. There is currently an ongoing large-scale project to sequence the genomes of seven dermatophyte species [26•], with the data available from the Broad Institute web site ([www.tinyurl.com/dermacomp7](http://www.tinyurl.com/dermacomp7)). Burmester et al. [27••] recently reported the analyses of the genomes of two of these, *Arthroderma benhamiae* and *Trichophyton verrucosum*, which infect rodents and cattle respectively, and also cause disease in humans.

The two genomes were assembled using data from a combination of next generation and first generation sequencing technologies, and are highly co-linear with only five major rearrangements. Although approximately 97% of the genes have clear orthologs in both species, there are >200 open reading frames that are specific to only one and may be important for host-specificity.

The genomes of the two dermatophytes were compared to the closely related species *Coccidioides posadasii* and *Coccidioides immitis*, and the more distant relative *Aspergillus fumigatus*. All these five dimorphic fungi contain substantial numbers of protease genes, but the dermatophytes are particularly protease-rich, with 235 putative proteases. In particular, the M35 (deuterolysin), M36 (fungolysin) and S8 (subtilisin) protease families are expanded in the dermatophytes, and several are

secreted during growth on keratin. The protease expansion is likely to reflect the ability of these species to use keratin as sole source of carbon and nitrogen.

The dermatophytes also have a different distribution of secondary metabolism gene clusters compared to their close relatives. Both species have a higher number of gene clusters than *Coccidioides*, mostly due to an increase in the number of non-ribosomal peptide synthetase (NRPS)-encoding genes. One NRPS and an associated MFS drug transporter are specific to *A. benhamiae*, although it is unknown whether this affects virulence and/or host specificity. There are also differences in the type of polyketide synthase (PKS) genes; most PKS are reducing in the dermatophytes, whereas the majority are non-reducing in other Ascomycetes, including *Coccidioides* species. The reducing PKS may be required for pigment formation, which is likely to be associated with virulence.

Abadio et al. [14] illustrated how the increasing availability of genome sequences might be used to identify new targets for antifungal drugs. They selected 55 genes that were shown experimentally to be essential in either *C. albicans* [28] or *A. fumigatus* [29] plus two non-essential genes that are important for cell viability within the host, *KRE2* and *ERG6* [30, 31]. Ten potential targets were selected among genes with orthologs in 8 fungal pathogens (*C. albicans*, *Aspergillus fumigatus*, *Cryptococcus neoformans*, *Paracoccidioides lutzii*, *Paracoccidioides brasiliensis*, *Blastomyces dermatitidis*, *Coccidioides immitis* and *Histoplasma capsulatum*) and absent from the human genome. These were reduced to four (*TRR1*, *RIM8*, *ERG6* and *KRE2*) using other criteria such as enzymatic activity, no auxotrophic phenotype, and accessibility in the cell. *TRR1* is required for maintaining redox status in *S. cerevisiae* [32] and *RIM8* is involved in a pathway regulating alkaline pH response [33] and in the activation of the yeast-hyphal morphological switch in *C. albicans* [34]. *ERG6*

encodes a sterol methyltransferase required for membrane permeability in *C. albicans* [31], whereas *KRE2* encodes a mannosyltransferase and is involved in N-linked glycosylation, cell adherence and virulence [35]. Homology modeling was used to predict the 3D structure of Trr1 and Kre2. The usefulness of this approach and its potential to reduce the huge costs of drug development [36] will become clearer in the future as the targets are virtually tested using chemical libraries.

Fungal genomes of pathogenic and non-pathogenic species will continue to be sequenced at a high rate, driven in particular by the 1000 fungal genomes project (<http://1000.fungalgenomes.org/home/>). However, it is also likely that sequencing multiple isolates of a single species will be used more frequently to yield biological insights. Such approaches have been used to study genome variation, diversity and evolution in *Saccharomyces* and related species [37, 38].

## **RNA-seq analysis of pathogenic fungi**

RNA-seq is still a new technology, and is slowly being applied to analysis of human fungal pathogens. It is generally used as a method for improving genome annotation, and is increasingly applied to analysis of differential gene expression. RNA-seq is likely to become the method of choice for transcriptome analysis in the next few years. We describe here some of the recent applications in *Candida*, *Aspergillus* and *Cryptococcus* species.

### ***Candida albicans***

RNA-seq analysis was first used to characterize the transcriptome of *C. albicans* under a variety of growth conditions, including hyphal induction, pH stress, oxidative stress, nitrosative stress and induction of cell wall-damage [39•, 40•]. The analyses

helped update the existing annotation and led to the characterization of 602 novel transcriptionally active regions (nTARS) that may represent non-coding or regulatory RNAs.

Nobile et al. [41••] subsequently used RNA-seq together with microarray analysis to further expand the identification of nTARs and to characterize the regulatory network required for biofilm formation. Biofilms are a community of microorganisms that grow on solid or liquid surfaces, such as indwelling medical devices, or host tissues. In *C. albicans*, biofilm formation is associated with a switch from predominantly yeast-like growth to hyphal or filamentous growth, followed by the production of an extracellular matrix [42-44]. Nobile et al. [41••] identified six transcription factors (*BCR1*, *TEC1*, *EFG1*, *NDT80*, *ROB1* and *BRG1*) that are required for biofilm development, both *in vitro* and *in vivo*, using established rat models for catheter [45] and oral infection [46]. Using a combination of transcriptional profiling and chromatin immunoprecipitation (ChIP) experiments they showed that the transcription factors and their targets belong to a single network that regulates the formation of biofilms. Each of the six regulators is a target of at least two of the others, and each also positively regulates its own transcription: the core network is therefore very tightly regulated. Their analysis indicated that the network evolved recently and is specific to *C. albicans*.

In contrast, Tierney et al. [47••] used RNA-seq to profile the response of both host and pathogen to infection. They followed the interaction of *C. albicans* with murine bone-marrow derived cells (BMDCs) over a 2 hour time course, from initial fungal adhesion to early lysis of host cells. Over the entire time course, 545 genes from *C. albicans* and 240 genes from *Mus musculus* were differentially expressed, and were placed into clusters based on common transcriptional patterns. Five *C.*



*albicans* genes and six *M. musculus* genes with roles in virulence or infection were selected to represent the clusters. A network model inferred from the data suggested that there are multiple interactions between *C. albicans* and mouse genes, including a hub represented by *HAP3*, a *C. albicans* transcription factor. The model predicted that *HAP3* was repressed by murine *PTX3*, and that *HAP3* in turn represses the mouse *MTA2* gene. *PTX3* encodes a pattern recognition receptor that has been shown to have an opsonin-like role in facilitating pathogen uptake by phagocytes [48]. In contrast, *MTA2* is associated with the NuRD chromatin remodeling machine [49] and knockouts display embryonic lethality with severe developmental issues in survivors, including autoimmune disease [50]. The predicted interaction between fungal *HAP3* and murine *PTX3* was confirmed experimentally by showing that *PTX3* is induced in bone marrow-derived macrophages (BMDMs) at both the protein and RNA level following infection with *C. albicans*, unless *HAP3* is deleted. Expression of murine *PTX3* is not induced by *C. albicans* cells carrying a knockout in *CDA2*, a putative chitin deacetylase with a binding motif for Hap3. Tierney et al. [47••] suggest that binding of Ptx3 to *Candida* cells alters the immune response of the host cells in a Hap3-dependent manner, which may involve a direct interaction between Hap3 in the fungus and Mta2 in the host.

This analysis illustrated the potential power of using RNA-seq to explore pathogen-host interactions; however, the approach is also limited by the size of the dataset, which can be overwhelming. Tierney et al. [47••] restricted their analyses to a small number of genes supported by previous experimental evidence, and focused on transcription factor “hubs”. While a similar approach has been used to identify networks associated with drug resistance in *Candida* [51], this method is only

successful when previous experimental analyses are available and emphasizes the need for supporting laboratory-based research.

### ***Candida parapsilosis***

The genome of *C. parapsilosis* was first sequenced in 2009 [10] together with several other pathogenic *Candida* species. In 2011, Guida et al. [52•] used RNA-seq to greatly improve the genome annotation and to characterize the transcriptional profile of *C. parapsilosis* across a variety of conditions, including temperature variations and oxygen levels. In total, 421 gene models were added to the previous annotation [10], 318 were removed and 664 were edited. Changes included the identification of 422 introns in 387 genes. Two introns were identified in the 3' UTR of the same gene, CPAR2\_601470; these are the first 3' introns described in *Candida* species, and suggest that a novel mechanism for regulating gene expression may be present. The RNA-seq data also provided evidence for at least 95 nTARs, suggesting that these are common in *Candida* species [39•].

Guida et al. also used RNA-seq and microarray analysis to determine the transcriptional response of *C. parapsilosis* to hypoxic (low oxygen) conditions, such as those encountered during infection or in biofilms [53, 54]. Both technologies identified increased expression of genes involved in ergosterol metabolism, fatty acid synthesis and glycolysis, and decreased expression of respiratory genes. The overall response is conserved in most fungi [53]. The analysis also identified the transcription factor *UPC2* as a major regulator of ergosterol synthesis and drug resistance genes. This role is conserved with *C. albicans* and *S. cerevisiae* [55, 56].

### ***Aspergillus***

Among the several hundred *Aspergillus* species identified, *Aspergillus flavus* and *Aspergillus fumigatus* are most closely associated with human infection [57, 58]. *A. flavus* produces several mycotoxins, including aflatoxins, which are an important contaminant of many food crops, such as corn, cotton and peanuts [59]. Aflatoxins are also highly carcinogenic, and ingestion is associated with acute hepatic failure [60]. Genes involved in aflatoxin biosynthesis are found close together in the genome in secondary metabolite clusters, including PKS and NRPS genes [61]. Yu et al. [62] used RNA-seq to determine the transcriptional response of 55 secondary metabolite clusters in *A. flavus* to changes in temperature. Expression of 11 clusters, including the aflatoxin cluster, is induced at 30°C relative to 37°C. The authors showed that most genes in the aflatoxin cluster are expressed at very low levels at the higher temperature. However, the sensitivity of the technology revealed that the expression levels of the two transcriptional regulators of the cluster (*aflR* and *aflS*) are 5-24 times higher at 30°C, although previous microarray experiments had suggested that they were equally expressed at both temperatures [63]. It is therefore possible that down regulation of the cluster at 37°C may be associated with low levels of expression of *aflS*.

Gibbons et al. [64•] applied RNA-seq to investigation of global transcription profiles of *A. fumigatus*. The authors compared the profiles of cells growing in liquid shake flasks (planktonic colonies, PL) to those growing in aerial static conditions, where they form a biofilm (BF) surrounded by an extracellular matrix [65]. The authors found that 96% of the 9887 annotated genes were expressed in at least one condition; only 251 genes were uniquely expressed in BF, and only 175 in PL cultures. However, the differentially expressed genes are not randomly located throughout the genome but are significantly enriched towards the telomeres.

Subsequent investigation revealed that these overlapped with secondary metabolite clusters, many of which were unregulated in biofilm conditions. A similar telomeric distribution of differentially expressed genes was previously observed during an analysis of early stages of infection of mammalian hosts with *A. fumigatus*. Genes involved in cell wall metabolism, ergosterol synthesis, multidrug transport and glycosylation are also upregulated in BF cells, whereas glycolytic genes are downregulated. The gene expression changes are likely to be associated with dramatic changes in the cell wall, which may contribute to the drug resistant phenotype of biofilm cultures.

RNA-seq has also been applied to transcriptional profiling of non-pathogenic *Aspergillus* species, such as *Aspergillus oryzae*, a species historically associated with the Japanese food industry, whose genome was sequenced in 2005 [66]. Similarly to the analysis described above for *C. parapsilosis*, Wang et al. [67] used RNA-seq to greatly improve the annotation of the *A. oryzae* genome. They reported widespread alternate splicing, in about 8.5% of all genes. Comparing the expression patterns of cells growing on solid agar compared to liquid cultures suggested that protein production is increased in cells grown on solid media. The authors also induced the unfolded protein response (UPR) by treatment with DTT, commonly activated by accumulation of unfolded proteins in the ER. Inducing UPR resulted in a significantly higher number of upregulated genes in cultures grown on solid media than in liquid cultures, suggesting that the efficiency of the response depends on the culture conditions.

### ***Cryptococcus neoformans***

Cryptococcus species are members of the Basidiomycete clade. Infection occurs through inhalation of spores, and cryptococcosis is particularly common in immunocompromised patients [68]. Infection may also spread to the brain where it causes cryptococcal meningitis [69]. The most common causative species are *Cryptococcus neoformans* and *Cryptococcus gatii*. One distinctive feature of these organisms is the presence of a polysaccharide capsule formed of glucuronoxylomannans (GXM) [70], which encapsulates the cell wall and is required for sexual development [71] and virulence [68, 72]. The capsule undergoes dramatic changes in size, particularly during infection [73].

Haynes et al. [74••] used RNA-seq together with ChIP-seq to characterize the transcriptional changes that occur during capsule growth. Several known signaling pathways had previously been linked to capsule formation, including the stress-induced cAMP pathway, regulation of iron sensing and the HOG pathway, although no integrated model had been proposed. Haynes et al. [74••] first identified genes whose transcriptional level correlated with capsule size using microarray technology, and selected *ADA2*, proposed to be part of the SAGA histone acetylation complex [75] for subsequent investigation. They showed that Ada2 is required for capsule development and for virulence in a mouse model. To further characterize the role of *ADA2*, the gene expression of the wild type and *ada2* mutant in capsule inducing and non-inducing conditions was determined using RNA-seq. Several members of the pheromone response pathway were downregulated in the *ada2* mutant, consistently with the role of capsule growth in mating. Genes related to capsule growth were also downregulated in both growth conditions.

RNA-seq was also used to determine the transcriptional profile of deletions of *CIR1* and *NRG1*, which results in hypocapsular formation. The two transcriptional

regulators also control sexual development [76]. Several genes required for capsule biosynthesis and for capsule regulation lie downstream of all three regulators.

To identify genes directly regulated by *ADA2*, the RNA-seq data was integrated with ChIP-seq analysis using antibodies specific for acetylated lysines (K9) on histone H3. Deleting *ada2* results in an overall reduction in the level of H3K9 acetylation, particularly near the transcriptional start sites of genes regulated by Ada2. These include *HXT1*, *CPL1* and *UGT1*, with known roles in capsule formation, *GAT201*, a positive regulator of capsule formation from the GATA family of transcription factors, and *STE2*, which encodes a mating pheromone receptor. Overall, the effects of *ADA2* are mostly indirect, as only 3% of genes have both differential expression and acetylation in an *ada2* deletion. It is likely that much of the response depends on 8 transcription factors that are directly regulated by ADA2. The integrated approach taken by Haynes et al. [74••] illustrates the power and resolution of next generation sequencing technologies.

### ***Cryptococcus gattii***

Unlike *C. neoformans*, *C. gattii* has only recently emerged as a pathogen in immunocompetent hosts. A well characterized outbreak began on Vancouver Island, Canada and subsequently spread to the Pacific [77]. Differences in virulence with respect to *C. neoformans* include an increased abundance of nodules (cryptococcomas) in lungs and in the brain and slower response to treatment [78].

At least four distinct lineages of *C. gattii* have been identified [79]; however, genome sequence is available for only two strains from two of the four lineages, isolated in Canada and Australia. To allow for more comprehensive analyses, Chaturvedi and Nierman [80] propose to sequence the genome and transcriptome of

twelve isolates. The selected strains cover all four lineages from a wide geographical distribution, including Africa, India, South and North America, and have been selected to provide a variety of suitable models of virulence as displayed in mice, rats and human.

Overall, the availability of these datasets would greatly enhance the ability to carry out comparative genomics analyses of *Cryptococcus* species, to identify targets for potential drugs and vaccines and to identify single nucleotide polymorphisms (SNPS) for molecular typing.

## Conclusion

Next-generation sequencing technology has led to an explosion in the amount of genome data available; for example, in a single year the Sequence Read Archive grew larger than the previous 20 years of data in the older GenBank repository [81••]. Several hundred fungal genomes have been completed or are in progress (<http://fungalgeneome.org>), and the number of RNA-seq data sets is increasing. The most commonly used methodologies are based on large-scale pyrosequencing (454 Life Sciences), ligation-based sequencing with di-base labeled probes (SOLiD) and sequencing by synthesis (Illumina) [82], with others currently under development [81••]. One of the drawbacks of RNA-seq is that data analysis can still be considered to be in a transitional phase. Even Illumina technology, benchmarked as the leading technology in 2010 [83], lacks a standard method of analysis, and a standard method of normalization. The most common strategy is to use RPKM or FPKM (Reads (or Fragments) per Kilobase of exon per Million mapped sequence reads), introduced in 2008 by Mortazavi et al. [2]. More recent developments include the introduction of geometric mean normalization [4] and quantile normalization [84].

In the last few years several open-source bioinformatic tools have been created to study differential gene expression through the quantification of RNA-seq data, such as the Tuxedo pipeline (including Cufflinks) [7••] and several R/Bioconductor packages like DESeq [4], baySeq [85], DEGseq [86] and edgeR [87]. Some of these freely available tools have been integrated in the Galaxy project, an online environment designed to be used by experimental researchers with little computational background [5]. For the fungal genomics studies we described here, no two groups used the same analysis strategy (Table 1). One of the priorities for the near future will be the emergence of a recognized standard, therefore increasing the reliability and our understanding of the vast amounts of information generated by next-generation sequencing.

Despite these outstanding issues, next-generation sequencing already made important contributions to our understanding of fungal virulence. The continued application of RNA-seq to fungal species in parallel with sequencing of new genomes and new strains will greatly increase our understanding of the physiology of the pathogens, as well as the interaction between pathogens and host.

## **Disclosure**

The authors report no potential conflict of interest relevant to this article.

## **References**

Papers of particular interest, published recently, have been highlighted as:

- Of importance
- Of major importance

1. Nagalakshmi U, Wang Z, Waern K, et al.: The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science*. 2008, 320:1344-9.



2. Mortazavi A, Williams BA, McCue K, et al.: Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods*. 2008;5:621-8.
3. Parkhomchuk D, Borodina T, Amstislavskiy V, et al.: Transcriptome analysis by strand-specific sequencing of complementary DNA. *Nucleic Acids Res*. 2009;37:e123.
4. Anders S, Huber W.: Differential expression analysis for sequence count data. *Genome Biol*. 2010;11:R106.
5. Blankenberg D, Gordon A, Von Kuster G, et al.: Manipulation of FASTQ data with Galaxy. *Bioinformatics*. 2010; 26:1783-5.
6. Wang X, Wu Z, Zhang X.: Isoform abundance inference provides a more accurate estimation of gene expression levels in RNA-seq. *J Bioinform Comput Biol*. 2010; 8 Suppl 1:177-92.
7. •• Trapnell C, Roberts A, Goff L, et al.: Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc*. 2012;7:562-78. *Description of the Tuxedo pipeline and its application for de novo transcriptome assembly, differential gene expression and visualization of results.*
8. Goffeau A, Barrell BG, Bussey H, et al.: Life with 6000 genes. *Science*. 1996;274:546-67.
9. Jones T, Federspiel NA, Chibana H, et al.: The diploid genome sequence of *Candida albicans*. *Proc Natl Acad Sci U S A*. 2004;101:7329-34.
10. Butler G, Rasmussen MD, Lin MF, et al.: Evolution of pathogenicity and sexual reproduction in eight *Candida* genomes. *Nature*. 2009;459:657-62.

11. Nierman WC, Pain A, Anderson MJ, et al.: Genomic sequence of the pathogenic and allergenic filamentous fungus *Aspergillus fumigatus*. *Nature*. 2005;438:1151-6.
12. Pel HJ, de Winde JH, Archer DB, et al.: Genome sequencing and analysis of the versatile cell factory *Aspergillus niger*. CBS 513.88. *Nat Biotechnol*. 2007;25:221-31.
13. Loftus BJ, Fung E, Roncaglia P, et al.: The genome of the basidiomycetous yeast and human pathogen *Cryptococcus neoformans*. *Science*. 2005;307:1321-4.
14. Abadio AK, Kioshima ES, Teixeira MM, et al.: Comparative genomics allowed the identification of drug targets against human fungal pathogens. *BMC Genomics*. 2011;12:75.
15. Cheng SC, Joosten LA, Kullberg BJ, et al.: Interplay between *Candida albicans* and the mammalian innate host defense. *Infect Immun*. 2012;80:1304-13.
16. • Silva S, Negri M, Henriques M, et al.: *Candida glabrata*, *Candida parapsilosis* and *Candida tropicalis*: biology, epidemiology, pathogenicity and antifungal resistance. *FEMS Microbiol Rev*. 2012;36:2880-305. *Review of biology of Candida albicans, Candida parapsilosis and Candida glabrata, including recent information about epidemiology and methods for laboratory identification*.
17. Jackson AP, Gamble JA, Yeomans T, et al.: Comparative genomics of the fungal pathogens *Candida dubliniensis* and *Candida albicans*. *Genome Res*. 2009;19:2231-44.

18. • Riccombeni A, Vidanes G, Proux-Wéra E, et al.: Sequence and analysis of the genome of the pathogenic yeast *Candida orthopsilosis*. PLoS One. 2012;7:e35750. *Recent example of hybrid de novo genome assembly, combining Sanger, 454 and Illumina sequencing technologies.*
19. Pfaller MA, Diekema DJ.: Epidemiology of invasive mycoses in North America. Crit Rev Microbiol. 2010;36:1-53.
20. Tavanti A, Davidson AD, Gow NA, et al.: *Candida orthopsilosis* and *Candida metapsilosis* spp. nov. to replace *Candida parapsilosis* groups II and III. J Clin Microbiol. 2005;43:284-92.
21. Gomez-Lopez A, Alastruey-Izquierdo A, Rodriguez D, et al.: Prevalence and susceptibility profile of *Candida metapsilosis* and *Candida orthopsilosis*: results from population-based surveillance of candidemia in Spain. Antimicrob Agents Chemother. 2008;52:1506-9.
22. Tay ST, Na SL, Chong J.: Molecular differentiation and antifungal susceptibilities of *Candida parapsilosis* isolated from patients with bloodstream infections. J Med Microbiol. 2009;58:185-91.
23. Zordan RE, Galgoczy DJ, Johnson AD.: Epigenetic properties of white-opaque switching in *Candida albicans* are based on a self-sustaining transcriptional feedback loop. Proc Natl Acad Sci U S A. 2006;103:12807-12.
24. Zhang A, Petrov KO, Hyun ER, et al.: The Tlo proteins are stoichiometric components of *Candida albicans* mediator anchored via the Med3 subunit. Eukaryot Cell. 2012;11:874-84.
25. Silva AP, Miranda IM, Guida A, et al.: Transcriptional profiling of azole-resistant *Candida parapsilosis* strains. Antimicrob Agents Chemother. 2011;55:3546-56.

26. • Achterman RR, White TC.: A foot in the door for dermatophyte research. PLoS Pathog. 2012;8:e1002564. *Description of ongoing experiments in genome sequencing and transcriptional profiling of dermatophyte species.*
27. •• Burmester A, Shelest E, Glöckner G, et al.: Comparative and functional genomics provide insights into the pathogenicity of dermatophytic fungi. Genome Biol. 2011;12:R7. *Example of integration of RNA-seq, secretome analysis and comparative genomics in the analysis of virulence in dermatophytes.*
28. Roemer T, Jiang B, Davison J, et al.: Large-scale essential gene identification in *Candida albicans* and applications to antifungal drug discovery. Mol Microbiol. 2003;50:167-81.
29. Hu W, Sillaots S, Lemieux S, et al.: Essential gene identification and drug target prioritization in *Aspergillus fumigatus*. PLoS Pathog. 2007;3:e24.
30. Buurman ET, Westwater C, Hube B, et al.: Molecular analysis of CaMnt1p, a mannosyl transferase important for adhesion and virulence of *Candida albicans*. Proc Natl Acad Sci U S A. 1998;95:7670-5.
31. Jensen-Pergakes KL, Kennedy MA, Lees ND, et al.: Sequencing, disruption, and characterization of the *Candida albicans* sterol methyltransferase (*ERG6*) gene: drug susceptibility studies in *erg6* mutants. Antimicrob Agents Chemother. 1998;42:1160-7.
32. Williams CH, Arscott LD, Müller S, et al.: Thioredoxin reductase two modes of catalysis have evolved. Eur J Biochem. 2000;267:6110-7.
33. Peñalva MA, Arst HN Jr.: Regulation of gene expression by ambient pH in filamentous fungi and yeasts. Microbiol Mol Biol Rev. 2002;66:426-46.

34. Davis D, Edwards JE Jr, Mitchell AP, Ibrahim AS: *Candida albicans* RIM101 pH response pathway is required for host-pathogen interactions. Infect Immun. 2000;10:5953-9.
35. Wagener J, Echtenacher B, Rohde M, et al.: The putative alpha-1,2-mannosyltransferase AfMnt1 of the opportunistic fungal pathogen *Aspergillus fumigatus* is required for cell wall stability and full virulence. Eukaryot Cell. 2008;7:1661-73.
36. Paul SM, Mytelka DS, Dunwiddie CT, et al.: How to improve R&D productivity: the pharmaceutical industry's grand challenge. Nat Rev Drug Discov. 2010;9:203-14.
37. Louis E.: *Saccharomyces cerevisiae*: gene annotation and genome variability, state of the art through comparative genomics. Methods Mol Biol. 2011;759:31-40.
38. Liti G, Schacherer J.: The rise of yeast population genomics. C R Biol. 2011;334:612-9.
39. • Bruno VM, Wang Z, Marjani SL, et al.: Comprehensive annotation of the transcriptome of the human fungal pathogen *Candida albicans* using RNA-seq. Genome Res. 2010;20:1451-8. *Example of the application of strand-specific RNA-seq.*
40. • Tuch BB, Mitrovich QM, Homann OR, et al.: The transcriptomes of two heritable cell types illuminate the circuit governing their differentiation. PLoS Genet. 2010;6:e1001070. *Characterization of an epigenetic switch in C. albicans using RNA-seq, and identification of nTARs.*
41. •• Nobile CJ, Fox EP, Nett JE, et al.: A recently evolved transcriptional network controls biofilm development in *Candida albicans*. Cell.

- 2012;148:126-38. *Outstanding example of integrated analysis, combining animal models, microarrays, RNA-seq, ChIP/chip and computational molecular evolution to describe the core regulatory network for biofilm formation in C. albicans.*
42. Harriott MM, Noverr MC.: Importance of *Candida*-bacterial polymicrobial biofilms in disease. Trends Microbiol. 2011;19:557-63.
  43. Sudbery PE.: Growth of *Candida albicans* hyphae. Nat Rev Microbiol. 2011;9:737-48.
  44. Martin R, Wächtler B, Schaller M, et al.: Host-pathogen interactions and virulence-associated genes during *Candida albicans* oral infections. Int J Med Microbiol. 2011;301:417-22.
  45. Andes D, Nett J, Oschel P, et al.: Development and characterization of an *in vivo* central venous catheter *Candida albicans* biofilm model. Infect Immun. 2004;72:6023-31.
  46. Nett J, Andes D.: *Candida albicans* biofilm development, modeling a host-pathogen interaction. Curr Opin Microbiol. 2006;9:340-5.
  47. •• Tierney L, Linde J, Müller S, et al.: An interspecies regulatory network inferred from simultaneous RNA-seq of *Candida albicans* invading innate immune cells. Front Microbiol. 2012;3:85. *Example of using RNA-seq to profile both host and pathogen response to infection.*
  48. Diniz SN, Nomizo R, Cisalpino PS, et al.: *PTX3* function as an opsonin for the dectin-1-dependent internalization of zymosan by macrophages. J Leukoc Biol. 2004;75:649-56.
  49. Feng Q, Zhang Y: The NuRD complex: linking histone modification to nucleosome remodeling. Curr Top Microbiol Immunol. 2003, 274:269-90.

50. Lu X, Kovalev GI, Chang H, et al.: Inactivation of NuRD component Mta2 causes abnormal T cell activation and lupus-like autoimmune disease in mice. *J Biol Chem.* 2008;283:13825-33.
51. Altwasser R, Linde J, Buyko E, et al.: Genome-wide scale-free network inference for *Candida albicans*. *Front Microbiol.* 2012;3:51.
52. • Guida A, Lindstädt C, Maguire SL, et al.: Using RNA-seq to determine the transcriptional landscape and the hypoxic response of the pathogenic yeast *Candida parapsilosis*. *BMC Genomics.* 2011;12:628. *First report of RNA-seq analysis in C. parapsilosis, including characterization of the hypoxic response.*
53. Ernst JF, Tielker D.: Responses to hypoxia in fungal pathogens. *Cell Microbiol.* 2009;11:183-90.
54. Rossignol T, Ding C, Guida A, et al.: Correlation between biofilm formation and the hypoxic response in *Candida parapsilosis*. *Eukaryot Cell.* 2009;8:550-9.
55. Synnott JM, Guida A, Mulhern-Haughey S, et al.: Regulation of the hypoxic response in *Candida albicans*. *Eukaryot Cell.* 2010;9:1734-46.
56. Vik A, Rine J.: Upc2p and Ecm22p, dual regulators of sterol biosynthesis in *Saccharomyces cerevisiae*. *Mol Cell Biol.* 2001;21:6395-405.
57. Marr KA, Carter RA, Crippa F, et al.: Epidemiology and outcome of mould infections in hematopoietic stem cell transplant recipients. *Clin Infect Dis.* 2002;34:909-17.
58. Morgan J, Wannemuehler KA, Marr KA, et al.: Incidence of invasive aspergillosis following hematopoietic stem cell and solid organ

- transplantation: interim results of a prospective multicenter surveillance program. *Med Mycol.* 2005;43 Suppl 1:S49-58.
59. Bennett JW, Klich M.: Mycotoxins. *Clin Microbiol Rev.* 2003;16:497-516.
  60. Paterson RR, Lima N.: Toxicology of mycotoxins. *EXS.* 2010;100:31-63.
  61. Yu J, Chang PK, Ehrlich KC, et al.: Clustered pathway genes in aflatoxin biosynthesis. *Appl Environ Microbiol.* 2004;70:1253-62.
  62. Yu J, Fedorova ND, Montalbano BG, et al.: Tight control of mycotoxin biosynthesis gene expression in *Aspergillus flavus* by temperature as revealed by RNA-Seq. *FEMS Microbiol Lett.* 2011;322:145-9
  63. O'Brian GR, Georgianna DR, Wilkinson JR, et al.: The effect of elevated temperature on gene transcription and aflatoxin biosynthesis. *Mycologia.* 2007;99:232-9.
  64. • Gibbons JG, Beauvais A, Beau R, et al.: Global transcriptome changes underlying colony growth in the opportunistic human pathogen *Aspergillus fumigatus*. *Eukaryot Cell.* 2012;11:68-78. *Using RNA-seq to determine the transcriptional profile of Aspergillus biofilms.*
  65. Loussert C, Schmitt C, Prevost MC, et al.: *In vivo* biofilm composition of *Aspergillus fumigatus*. *Cell Microbiol.* 2010;12:405-10.
  66. Machida M, Asai K, Sano M, et al.: Genome sequencing and analysis of *Aspergillus oryzae*. *Nature.* 2005;438:1157-61.
  67. Wang B, Guo G, Wang C, et al.: Survey of the transcriptome of *Aspergillus oryzae* via massively parallel mRNA sequencing. *Nucleic Acids Res.* 2010;38:5075-87.



68. Garcia-Hermoso D, Janbon G, Dromer F.: Epidemiological evidence for dormant *Cryptococcus neoformans* infection. J Clin Microbiol. 1999;37:3204-9.
69. Brizendine KD, Baddley JW, Pappas PG.: Pulmonary cryptococcosis. Semin Respir Crit Care Med. 2011;32:727-34.
70. Martinez LR, Casadevall A.: *Cryptococcus neoformans* biofilm formation depends on surface support and carbon source and reduces fungal cell susceptibility to heat, cold, and UV light. Appl Environ Microbiol. 2007;73:4592-601.
71. Chang YC, Miller GF, Kwon-Chung KJ.: Importance of a developmentally regulated pheromone receptor of *Cryptococcus neoformans* for virulence. Infect Immun. 2003;71:4953-60.
72. Doering TL.: How sweet it is! Cell wall biogenesis and polysaccharide capsule formation in *Cryptococcus neoformans*. Annu Rev Microbiol. 2009;63:223-47.
73. Rivera J, Feldmesser M, Cammer M, Casadevall A.: Organ-dependent variation of capsule thickness in *Cryptococcus neoformans* during experimental murine infection. Infect Immun. 1998;66:5027-30.
74. •• Haynes BC, Skowrya ML, Spencer SJ, et al.: Toward an integrated model of capsule regulation in *Cryptococcus neoformans*. PLoS Pathog. 2011;7:e1002411. *An example of integrating RNA-seq with ChIP-seq to identify genes directly involved in capsule regulation.*
75. Koutelou E, Hirsch CL, Dent SY.: Multiple faces of the SAGA complex. Curr Opin Cell Biol. 2010;22:374-82.

76. Jung WH, Sham A, White R, et al.: Iron regulation of the major virulence factors in the AIDS-associated pathogen *Cryptococcus neoformans*. PLoS Biol. 2006;4:e410.
77. Carriconde F, Gilgado F, Arthur I, et al.: Clonality and  $\alpha$ -a recombination in the Australian *Cryptococcus gattii* VGII population--an emerging outbreak in Australia. PLoS One. 2011;6:e16936.
78. Galanis E, Hoang L, Kibsey P, et al.: Clinical presentation, diagnosis and management of *Cryptococcus gattii* cases: Lessons learned from British Columbia. Can J Infect Dis Med Microbiol. 2009, 20:23-8.
79. Bovers M, Hagen F, Boekhout T.: Diversity of the *Cryptococcus neoformans*-*Cryptococcus gattii* species complex. Rev Iberoam Micol. 2008;25:S4-12.
80. Chaturvedi V, Nierman WC.: *Cryptococcus gattii* comparative genomics and transcriptomics: a NIH/NIAID white paper. Mycopathologia. 2012;173:367-73.
81. •• Thompson JF, Milos PM.: The properties and applications of single-molecule DNA sequencing. Genome Biol. 2011;12:217. *Extensive review of next-generation sequencing, describing benefits, drawbacks and applications of the main technologies.*
82. Rizzo JM, Buck MJ.: Key principles and clinical applications of "next-generation" DNA sequencing. Cancer Prev Res (Phila). 2012;5:887-900.
83. Levin JZ, Yassour M, Adiconis X, et al.: Comprehensive comparative analysis of strand-specific RNA sequencing methods. Nat Methods. 2010;7:709-15.
84. Hansen KD, Irizarry RA, Wu Z.: Removing technical variability in RNA-seq data using conditional quantile normalization. Biostatistics. 2012;13:204-16.

85. Hardcastle TJ, Kelly KA.: baySeq: empirical bayesian methods for identifying differential expression in sequence count data. BMC Bioinformatics. 2010;11:422.
86. Wang L, Feng Z, Wang X, et al.: DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. Bioinformatics. 2010;26:136-8.
87. Robinson MD, McCarthy DJ, Smyth GK.: edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26:139-40.
88. Brown JA, Sherlock G, Myers CL, et al.: Global analysis of gene function in yeast by quantitative phenotypic profiling. Mol Syst Biol. 2006;2:2006.0001.
89. Li H, Ruan J, Durbin R.: Mapping short DNA sequencing reads and calling variants using mapping quality scores. Genome Res. 2008;18:1851-8.
90. Jiang H, Wong WH.: SeqMap: mapping massive amount of oligonucleotides to the genome. Bioinformatics. 2008;15:2395-6.
91. Jiang H, Wong WH.: Statistical inferences for isoform expression in RNA-Seq. Bioinformatics. 2009;25:1026-32.
92. Smyth GK.: Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol. 2004;3:Article3
93. Lo K, Gottardo R.: Flexible empirical bayes models for differential gene expression. Bioinformatics. 2007;23:328-35.
94. Li R, Li Y, Kristiansen K, Wang J.: SOAP: short oligonucleotide alignment program. Bioinformatics. 2008;24:713-4.

95. Marioni JC, Mason CE, Mane SM, et al.: RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res.* 2008;18:1509-17.
96. Kent WJ.: BLAT--the BLAST-like alignment tool. *Genome Res.* 2002;12:656-64.
97. Zhang M, Gish W.: Improved spliced alignment from an information theoretic approach. *Bioinformatics.* 2006;22:13-20.

**Table 1:** Comparison of tools used to analyze RNA-seq data from fungal datasets

Paper	Sequencing method	Organism	Reads aligner	Mapped count normalization	Differential gene expression
Bruno et al. 2010 [39••]	Illumina	<i>C. albicans</i>	TopHat [7••]	RPKM (custom)	Hierarchical clustering [88]
Gibbons et al. 2011 [64•]	Illumina	<i>A. fumigatus</i>	Maq [89], SeqMap [90]	RPKM, rSeq [91]	RPKM ratio
Guida et al. 2011 [52]	Illumina	<i>C. parapsilosis</i>	TopHat	RPKM, quartile normalization	Cufflinks
Haynes et al. 2011 [74••]	Illumina	<i>C. neoformans</i>	TopHat	RPKM, quartile normalization	Cufflinks, LIMMA [92], ELNN [93]
Yu et al. 2011 [62]	Illumina	<i>A. flavus</i>	CLC Genomics Workbench (clcbio.com)	RPKM (custom)	CLC Genomics Workbench
Tierney et al. 2012 [47••]	Illumina	<i>C. albicans</i> , <i>M. musculus</i>	TopHat	RPKM	baySeq [85]
Wang et al. 2012 [67]	Illumina	<i>A. oryzae</i>	SOAP [94]	RPKM	DEGseq [86]
Nobile et al. 2012 [41••]	SOLiD	<i>C. albicans</i>	Life Technologies Whole Transcriptome Software Tools [40•] (lifetechnologies.com)	pseudo-RPKM	Likelihood ratio [95]
Burmester et al. 2011 [27••]	454	<i>A. benhamiae</i> , <i>T. verrucosum</i>	BLAT [96], Exalin [97]	raw counts	DESeq [4], edgeR [87]