Journal of Ornithology (2018) 159:599–629
https://doi.org/10.1007/s10336-018-1532-5

REVIEW

Avian transcriptomics: opportunities and challenges

Elinor Jax[1,2] · **Michael Wink**[3] · **Robert H. S. Kraus**[1,2]

Received: 29 September 2017 / Revised: 27 December 2017 / Accepted: 15 January 2018 / Published online: 5 February 2018

Abstract

Recent developments in next-generation sequencing technologies have greatly facilitated the study of whole transcriptomes in model and non-model species. Studying the transcriptome and how it changes across a variety of biological conditions has had major implications for our understanding of how the genome is regulated in different contexts, and how to interpret adaptations and the phenotype of an organism. The aim of this review is to highlight the potential of these new technologies for the study of avian transcriptomics, and to summarise how transcriptomics has been applied in ornithology. A total of 81 peer-reviewed scientific articles that used transcriptomics to answer questions within a broad range of study areas in birds are used as examples throughout the review. We further provide a quick guide to highlight the most important points which need to be take into account when planning a transcriptomic study in birds, and discuss how researchers with little background in molecular biology can avoid potential pitfalls. Suggestions for further reading are supplied throughout. We also discuss possible future developments in the technology platforms used for ribonucleic acid sequencing. By summarising how these novel technologies can be used to answer questions that have long been asked by ornithologists, we hope to bridge the gap between traditional ornithology and genomics, and to stimulate more interdisciplinary research.

Keywords Next-generation sequencing · Genetics · Ornithology · Gene expression · RNA-sequencing

Zusammenfassung

Vogeltranskriptomik – Chancen und Herausforderungen

Erst die Entwicklungen auf dem Gebiet der next-generation sequencing (NGS) Technologien in den letzten 10 Jahren haben die Untersuchung kompletter Transkriptome in Modell- sowie nicht-Modell-Organismen ermöglicht. Die Untersuchung des Transkriptoms und seiner Regulation hatte maßgeblichen Einfluss auf unser Verständnis darüber, welche Gene in welchen Geweben und zu welchem Zeitpunkt aktiv sind, und wie wir heute Anpassungen und generell den Phänotyp eines Organismus verstehen. Das Ziel unseres Übersichtsartikels besteht darin, das Potential der neuen Technologien zum Studium der Transkriptomik bei Vögeln herauszustellen. Wir erläutern, wie sie bisher schon in der Ornithologie, die sich überwiegend mit nicht-Modell-Organismen beschäftigt, Anwendung gefunden haben. Insgesamt 81 wissenschaftlich begutachtete Transkriptomik Fachartikel, die breitgefächerte Fragen der Vogelkunde untersuchen, werden von uns in diesem Übersichtsartikel als Beispiele vorgestellt. Weiterhin bieten wir dem Teil der ornithologischen Gemeinde, der wenig Hintergrundwissen in molekularen Methoden hat, eine kurze Einführung in die Planung eines Trankriptomikprojektes bei Vögeln. Die wichtigsten Punkte, die schon früh in der

Communicated by J. T. Lifjeld.

✉ Elinor Jax
ejax@orn.mpg.de

[1] Department of Migration and Immuno-Ecology, Max Planck Institute for Ornithology, Am Obstberg 1, 78315 Radolfzell, Germany

[2] Department of Biology, University of Konstanz, 78457 Konstanz, Germany

[3] Institute of Pharmacy and Molecular Biotechnology, Heidelberg University, INF 364, 69120 Heidelberg, Germany

Planung zu beachten sind, um mögliche Fehler zu vermeiden, werden diskutiert. Dies beinhaltet überall auch Vorschläge zur vertiefenden Lektüre weiterführender wissenschaftlicher Artikel. Zum Schluss diskutieren wir zukünftige Entwicklungen der verschiedenen Technologien und neue Plattformen, die bei RNA Sequenzierung zur Anwendung kommen könnten. Durch unsere breit angesetzte Zusammenfassung dieser neuen und zukünftigen Technologien zeigen wir auf, wie Ornithologen lang unbeantwortete Grundsatzfragen angehen können. Wir hoffen, dass wir dazu beitragen, eine Lücke zwischen der Molekularbiologie und Genomik auf der einen und traditioneller Ornithologie auf der anderen Seite schlagen zu können.

Background

The genetic information of an organism is stored in its genome. While the genomes of some viruses are made of the molecule ribonucleic acid (RNA), the majority of genomes are comprised of deoxyribonucleic acid (DNA). In multicellular animals, DNA is stored in chromosomes (in the nucleus; ncDNA) and in organelles (such as mitochondria; mtDNA). Most vertebrates have genomes with roughly 20,000 protein-coding genes representing only 2–3% of the genome. The rest of the DNA includes ribosomal DNA (rDNA) genes, pseudogenes, DNA for non-coding RNA and repetitive DNA elements.

Whereas almost all cells in one individual multicellular organism carry an identical genome (known as the 'genotype'), every cell differs in the expression of its genes, leading to cell-, tissue- and development-specific phenotypes. For gene expression, information encoded in the genes is used for the synthesis of functional gene products. In the most well-known of these processes, DNA is transcribed into messenger RNA (mRNA), which is then translated into amino acid chains which are folded into functional proteins during protein synthesis (Fig. 1). Gene products can, however, also be functional RNA molecules, as is the case for ribosomal RNA (rRNA) and non-protein-coding genes (Wolf 2013). While non-coding DNA has previously been regarded as 'junk RNA', it is becoming evident that it has important regulatory roles in both normal cellular processes and disease states (Clancy 2008).

The complete set of RNA present in an organism, cell or tissue at a certain time point is called the 'transcriptome' of that organism, cell or tissue. In contrast to the genome, the transcriptome is highly dynamic throughout the life span of an organism. By measuring the content and composition of the mRNA in a biological sample, it is therefore possible to estimate which genes are being expressed at a particular time point in that sample. In addition, analysing the number of specific RNA molecules provides evidence of how strongly a gene is transcribed. Studying the transcriptome therefore gives us a better idea of which functions different coding or non-coding genes might have, and a better understanding of how they contribute to the phenotype of individuals and their functional adaptations during their lifetime.

Molecular biologists have been studying gene expression for some time. For many years, DNA microarray analyses were the methods of choice, but these were mainly available for model species. Recent developments of high-throughput or next-generation sequencing (NGS) technologies have greatly facilitated the study of transcriptomes, also in non-model species (so-called transcriptomics), and in the last few years there has been a rapid increase in the number of such studies in the scientific literature (Ekblom and Galindo 2011). The wide variety of applications of transcriptomics within fields ranging from developmental biology, molecular ecology and evolution to immunology and medicine has contributed to its increasing popularity (Alvarez et al. 2015; Todd et al. 2016).

In this review, we will highlight developments of transcriptome analyses in birds, starting from some of the first methodologies used to study gene expression until current high-throughput sequencing methodologies. With examples from birds, we discuss these technologies using study systems well known to the readers of the *Journal of Ornithology*, and also give those readers with little background in genomics an overview of the current state of avian transcriptomics. We will highlight the potentials of these new technologies for the study of avian transcriptomics, and discuss how to avoid potential pitfalls, thereby creating a resource comprising both conceptual summaries and case studies.

RNA analysis of birds before NGS: a brief history

A number of excellent reviews have already given a comprehensive overview of the history of methodologies developed for transcriptome analysis (Morozova et al. 2009; Schulze and Downward 2001; VanGuilder et al. 2008). Here, we will therefore only briefly mention some of the most commonly used pre-NGS RNA techniques that have been used to improve our understanding of gene expression patterns in non–model species (such as most birds), and we will briefly discuss the advantages and disadvantages of these techniques.

Real-time quantitative polymerase chain reaction

Real-time quantitative polymerase chain reaction (qPCR) is currently one of the most widely used methods for the study of gene expression of known candidate genes. When using

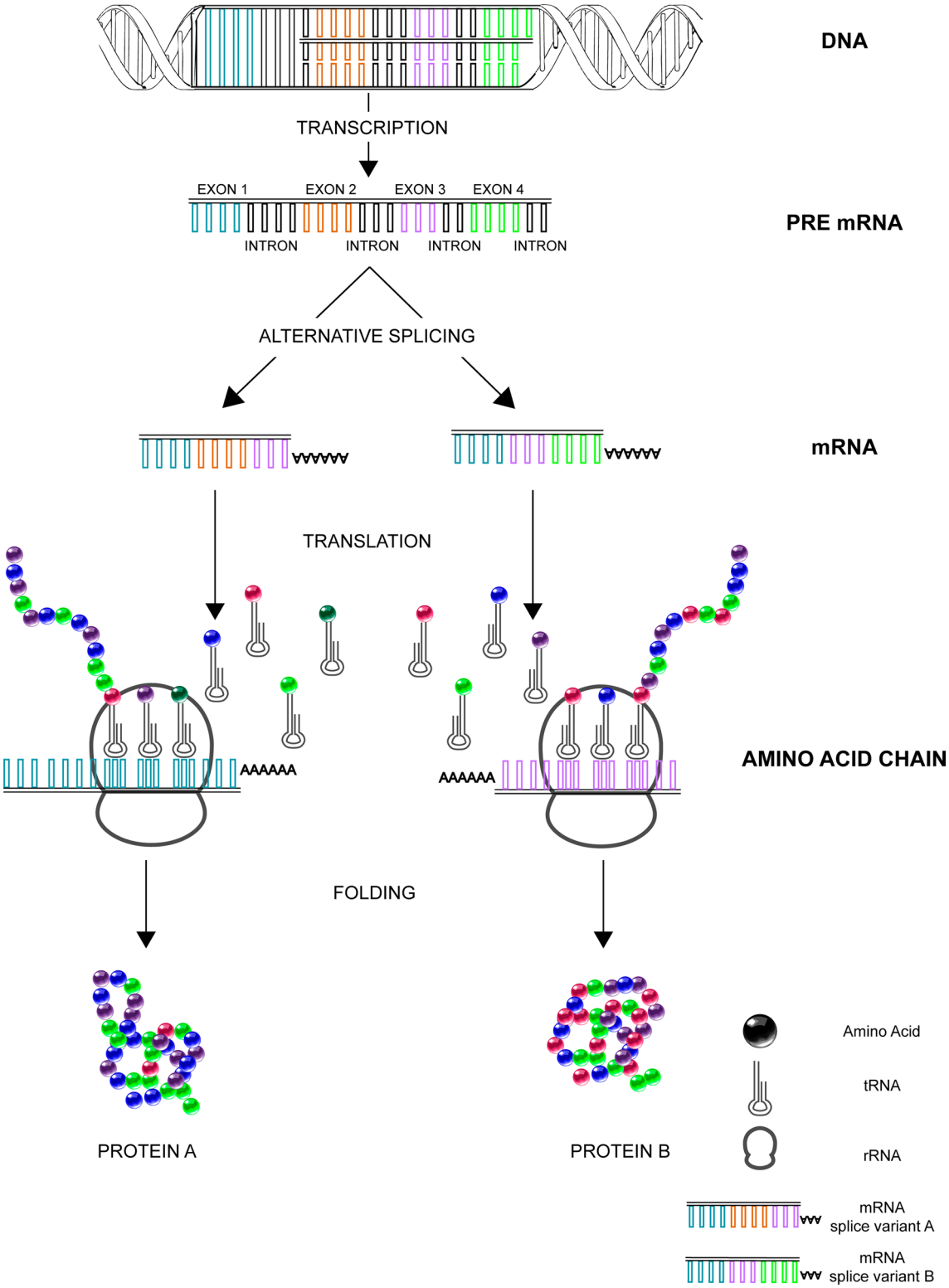

Fig. 1 Protein synthesis: protein-coding genes are transcribed into messenger ribonucleic acid (*mRNA*), which is consecutively translated into proteins in ribosomes. During this process, transfer RNAs (*tRNAs*) carry amino acids to the ribosome, which is composed of ribosomal RNA (*rRNA*) and ribosomal proteins. The ribosome is the structure in which protein synthesis takes place; the amino acids brought by the tRNA to the ribosome are connected to a growing amino acid chain, which in turn folds into a protein. In some cases, alternative splicing of pre-mRNA produces different isoforms of the same gene, e.g. in this example, protein A and protein B (for details see "Alternative splicing/transcriptome isoforms")

this method, complementary DNA (cDNA) is synthesised from isolated mRNA by reverse transcriptase. Gene-specific primers are then used to amplify the gene of interest, and the abundance of that particular cDNA molecule is monitored in real time during amplification using fluorescent dyes (Josefsen et al. 2012). From the kinetics of PCR product formation one can calculate the actual amount of starting RNA transcript and do quantitative comparisons between samples (Pfaffl 2012). However, the genes of interest must be known; this is often not the case when more complex phenomena are to be investigated. This technology has been used extensively to study gene expression in birds, for example with the aim of better understanding how bird species adapt to new environments (Martin et al. 2014), acclimatise to winter conditions (Swanson et al. 2009) or deal with infections (Fleming-Canepa et al. 2011). Furthermore, real-time qPCR has a number of applications other than the analysis of gene expression, such as transcript genotyping, virus detection or measuring the length of telomeres. Examples of this in birds include the genotyping of the major histocompatibility complex class I (Moon et al. 2005) and avian influenza virus detection (Fleming-Canepa et al. 2011; Kraus et al. 2009, 2011; Takekawa et al. 2011).

Real-time qPCR is one of the most accurate, sensitive and rapid techniques available for candidate gene expression studies (Derveaux et al. 2010). In comparison to many other RNA technologies, it is furthermore rather inexpensive and easy to undertake in-house; hence, it is available to most research groups. Concerns, however, have repeatedly been raised regarding (1) the lack of quality control and standardisation in a real-time qPCR workflow, and (2) the lack of sufficient details reported in many publications to enable qualitative evaluation of the results (Bustin et al. 2009; Chapman and Waldenström 2015; Derveaux et al. 2010; Nolan et al. 2006).

Recent advances allowing for parallelisation of real-time qPCR by microfluidic techniques now make this method attractive for studying gene expression in a large number of genes simultaneously (Spurgeon et al. 2008). Parallelisation technology has already been used to investigate gene expression in domestic chicken *Gallus gallus* (Van Goor et al. 2017) and Mule Duck *Anas platyrhynchos* × *Cairina moschata* (Annabelle et al. 2017). The necessity of prior knowledge of genes and their nucleotide sequences, however, limits the use of this technology in non-model species.

Microarray analysis

Microarray analysis is another technology that has been used in many model species, including birds. DNA microarrays have a number of applications, such as sequencing or measuring the expression levels of a large number of genes simultaneously (Harrington et al. 2000; Schena et al. 1995). DNA microarrays are small glass slides with a large number of DNA oligonucleotides (representing individual genes) bound to their surfaces. By fluorescent labelling of the sample of interest, hybridisation of the DNA (or cDNA) to its complementary sequences on the microarray can be detected and measured in parallel (Lockhart and Winzeler 2000). In gene expression studies, using different fluorescent labels for different samples enables the comparison of gene expression between groups and thus differential gene expression. Microarrays have been developed to study the expression of genes involved in a wide range of functions in avian model species, such as bird song and immune reactions (Afrakhte and Schultheiss 2004; Bliss et al. 2005; Burnside et al. 2005; Li et al. 2008; Neiman et al. 2001; Smith et al. 2006; Van Hemert et al. 2003; Wada et al. 2006; Wade et al. 2004). Microarrays customised for avian model species such as Zebra Finch *Taeniopygia guttata* (Naurin et al. 2008; Wada et al. 2006) have further been used for related species (termed 'cross-species microarray analysis') (Cheviron et al. 2008; Naurin et al. 2012).

Microarrays enable the study of hundreds of genes simultaneously, and have hence contributed to improving our understanding of the complexity of gene networks in birds. For those species where microarrays are available, this technology therefore offers great opportunities for the study of gene expression in a straightforward manner for a relatively low price. The necessity of prior information on a large number of genes, however, has limited the use of this technology for non-model species research. Further, when a DNA microarray developed e.g. for Zebra Finch is used to study gene expression of another bird species, it is likely that some transcripts will not be detected because the corresponding sequences differ between the taxa. As a consequence, some sequences of the target species will not bind to the microarray.

With the development and decreasing costs of NGS technologies, we therefore expect to see a shift away from using microarrays to NGS in non-model avian research. NGS further avoids the common issues of microarray technology, e.g. background hybridisation and cross-hybridisation of the cDNA and probes, which makes it difficult to detect rare transcripts as well as accurately measure expression levels of more common transcripts (Ekblom and Galindo 2011; Zhao et al. 2014).

The majority of the pre-NGS RNA analyses, including real-time qPCR and the most commonly used DNA microarrays, require prior knowledge of the genes of interest. Before the genomic era, these downstream applications were therefore limited to a small subset of genes in a few model species. As the field of avian genomics, and thereby the knowledge of the genetic code of a multitude of non-model

species, continues to progress (Kraus and Wink 2015) the potential use of these targeted downstream applications is likely to grow. Still, being able to study a number of preselected genes has several limitations, which can be avoided when directly using new NGS technologies.

Ornithology in the age of *omics

NGS, or 'massively parallel sequencing', refers to technologies that allow for the concomitant sequencing of millions of smaller pieces of DNA, so-called short-sequence reads (Chiu and Miller 2016). Recent advances in NGS technologies have made it possible to undertake in-depth transcriptomic and genomic analysis of ecologically important model species (Alvarez et al. 2015; Ellegren 2008; Tautz et al. 2010). These technologies have further enabled the ornithological community to study large-scale *omics in non-model species of birds (including genomics, transcriptomics, proteomics, metabolomics), thus providing a new tool to improve insight into the genetics and evolution of birds.

The first genomic studies of bird species were undertaken in model species with high importance in agriculture such as the chicken *G. gallus* (Hillier et al. 2004), Turkey *Meleagris gallopavo* (Dalloul et al. 2010) and Pekin Duck *Anas platyrhynchos domestica* (Huang et al. 2013). In 2010, the genome of the Zebra Finch *T. guttata*, an important model species for behaviour and neurobiology, was published (Warren et al. 2010). This was considered a big step forward in evolutionary and ecological studies in birds (Balakrishnan et al. 2010). Since then the genomic era of wild avian non-model species has taken off with, for example, studies on the Collared Flycatcher *Ficedula albicollis* (Ellegren et al. 2012), Peregrine Falcon *Falco peregrinus* and Saker Falcon *Falco cherrug* genomes (Zhan et al. 2013), followed by a *Science* issue on avian genomics with comparative studies on 48 avian genomes (Zhang et al. 2014a). Furthermore, in 2015, the Avian Phylogenomics Consortium launched the Bird 10,000 Genomes (B10K) project with the aim of generating draft genomes for all extant bird species within the next 5 years (Zhang 2015). Unfortunately, many of the newly published genomes are largely incomplete and not well annotated [although this will change, cf. Korlach et al. (2017)]. Nevertheless, the continuously growing database of avian genomes has important implications for downstream applications such as transcriptomics. For further reading on avian non-model genomics there are a number of recent reviews covering this topic (Kraus and Wink 2015; Oyler-McCance et al. 2016; Toews et al. 2015).

Recent advances in NGS technologies have further made it possible to characterise the whole RNA content and composition in a sample using whole-transcriptome shotgun sequencing (RNA-seq) (Wolf 2013). When using RNA-seq, information on the whole transcriptome can thus be acquired without prior knowledge of the transcripts in the target species (Wang et al. 2009). This method has revolutionised the study of transcriptomics in non-model species, and has resulted in a number of publications using RNA-seq to analyse the transcriptome in birds. The rapid developments of NGS technologies and analytical tools, as well as the decreasing costs for using these technologies, have also made them available to small research groups working in ornithology. In this review, we show some of the areas where RNA-seq has so far proven useful for ornithological studies of non-model species, and try to guide those scientists new to this technology through the workflow necessary when using this technology. In our opinion, RNA-seq is an important technology, which should be added to the toolbox to help us answer questions regarding the biology, development, ecology and behaviour of birds.

In the following section, we will briefly introduce the concepts of RNA-seq. For a more detailed description of the general RNA-seq workflow there are a number of informative reviews covering this topic in a non-taxon-specific manner (Bullard et al. 2010; Conesa et al. 2016; Mazzoni and Kadarmideen 2016; Oshlack et al. 2010; Vijay et al. 2013; Wang et al. 2009; Wolf 2013). In the "Planning an RNA-seq study: a quick guide" section of this review, we will further give a quick guide on what to take into consideration when planning an RNA-seq study in birds.

RNA-seq: a brief introduction

A typical RNA-seq workflow comprises field or laboratory work, wet laboratory processes, bioinformatics processing and biological inference (Wolf 2013). Briefly, after acquiring samples, the RNA of interest is isolated (for more details see "RNA isolation and library preparation"). The RNA is then converted to a fragmented cDNA library with molecular adaptors specific to the sequencing platform and the individual samples (Chu and Corey 2012). The library is then sequenced by NGS, resulting in a large number (millions) of short reads normally ranging from 50 to 300 nucleotides (nt) (although new technologies allow for sequencing the whole length of the transcripts that can span far more than 1000 nt; see "Choice of sequencing platform"). The quality of the sequencing reads is then controlled using bioinformatic tools, and low-quality sequence positions as well as duplicated reads or remaining adaptors are removed in a process called 'pre-processing' (Conesa et al. 2016). The remaining sequences are then either aligned to a reference genome or transcriptome (also called 'mapping'), or in the absence of

a reference genome assembled de novo (Oshlack et al. 2010; Vijay et al. 2013). For those interested in sequence diversity or genetic marker discovery, the assembled transcripts can then be used for further analysis (De Wit et al. 2012). In comparative transcriptomic studies the quantitative expression levels of the genes are used to infer differential expression between groups of interest. Here, the number of transcripts aligned to each gene is normalised, taking parameters such as gene length and total number of reads per sample into account (Bullard et al. 2010). For species without a reference genome, a gene identification step called 'gene annotation' might further be necessary (for more details see "Biological interpretation of RNA-seq data"). The last step of the RNA-seq workflow is the interpretation of the results in a biological context.

Applications of transcriptomics to non-model bird species

Two of the first RNA-seq studies undertaken in birds were published in 2010 in a special issue of *Molecular Ecology* 'Next Generation Molecular Ecology'. In one of the papers, Wolf et al. (2010) used transcriptomic data to study population differentiation and speciation in the hybridisation zone between the black Carrion Crow *Corvus* (*corone*) *corone* and the grey-coated Hooded Crow *Corvus* (*corone*) *cornix*. The authors found a clear clustering of the expression profiles when comparing the two morphs, and thus suggested that gene expression differences may be a sensitive indicator of initial species divergence (Wolf et al. 2010). In the same special issue, a second study analysed the pattern of mutation rate and selection across ten non-model avian species using brain transcriptomes, and among other findings the authors discovered a higher mutation rate of the Z chromosome than of autosomes (Künstner et al. 2010). This work was soon followed by a study investigating the tissue-specific gene expression profiles of the Zebra Finch (Ekblom et al. 2010a). By focusing on certain immune genes the authors showed that NGS data are useful for the analysis of gene expression of the whole genome as well as for candidate genes (Ekblom et al. 2010a). Since 2010, the number of transcriptome studies has increased in avian non-model species. In this section, we will give an overview of topics that can be studied using RNA-seq methodologies. We further illustrate each topic with a few specific examples from ornithological research, and present selected example studies in more detail in separate case study boxes. We provide a tabularisation (Table 1) in which currently available peer-reviewed scientific articles on this topic ($n = 81$) have been collected for future guidance of the interested reader. For the purpose of this section, *Google Scholar* was searched using the keywords 'transcriptomics', 'RNA-seq' or 'RNA sequencing' + 'avian' or 'bird' (up until 18 December 2017).

Characterisation of transcriptomes

While it may be straightforward to imagine what a reference genome is (Kraus and Wink 2015), a reference transcriptome needs more explanation. A reference transcriptome comprises a comprehensive set of transcripts that can be transcribed from the genome. As the transcriptome is cell, tissue and age specific, a reference transcriptome should preferably contain transcripts from a number of (or possibly all) different tissues and from different life stages. Having access to a high-quality reference transcriptome as well as a genome is a great advantage when undertaking transcriptomic studies, as these resources can be used to facilitate transcript assembly as well as gene identification (note, for more details see "Biological interpretation of RNA-seq data"). As of December 2017, there were only five well-annotated genomes of birds (i.e. those genomes annotated by extensive transcriptomics) available for download on the genomic data base Ensembl (www.ensembl.org). These species were chicken, Collared Flycatcher, Mallard, Turkey and Zebra Finch (Table 2).

Several avian transcriptomes have further been published as parts of specific studies in ecology and evolution with the aim of them serving as important resources for future transcriptomic studies (e.g. Chu et al. 2012; Ekblom et al. 2014; Koglin et al. 2017; Peterson et al. 2012; Richardson et al. 2017; Santure et al. 2011; Srivastava et al. 2012) (Box 1), sometimes in combination with the genome analyses of the same species (e.g. Chung et al. 2015; Mueller et al. 2015).

Table 1 Ribonucleic acid (RNA)-sequencing studies undertaken in non-model species of birds

Order	Species	Common name	Reference [research category] and {tissue}
Accipitriformes	*Aegypius monachus*	Cinereous Vulture	Chung et al. (2015) [1] {27}
	Aquila chrysaetos	Golden Eagle	Van Den Bussche et al. (2017) [1] {2, 15, 17}
Anseriformes	*Anas platyrhynchos*	Mallard/Pekin Duck	Huang et al. (2013) [1, 2] {2, 12, 16, 17, 21}; Wright et al. (2014) [3] {18, 22, 23}; Harrison et al. (2015) [3] {18, 22, 23}; Li et al. (2015) [2] {3, 22, 24}; Smith et al. (2015) [2] {12, 16}; Wright et al. (2015) [3] {18, 22, 23}
	Anser cygnoides	Swan Goose[a]	Harrison et al. (2015) [3] {18, 22, 23}; Lu et al. (2015) [1, 2] {15}; Tariq et al. (2015) [2] {28}; Wright et al. (2015) [3] {18, 22, 23}; Chen et al. (2017b) [2] {28}; Cao et al. (2017) [2] {44}
	Cairina moschata	Muscovy Duck	Wang et al. (2015b) [2] {22}; Wang et al. (2017b) [2] {15}
Apodiformes	*Aerodramus fuciphagus*	Edible-nest Swiftlet	Looi et al. (2017) [6] {38}
	Aerodramus maximus	Black-nest Swiftlet	Looi et al. (2017) [6] {38}
	Apus affinis	Little Swift	Looi et al. (2017) [6] {38}
	Archilochus colubris	Ruby-throated Hummingbird	Künstner et al. (2010) [4] {2}; Nabholz et al. (2011) [5] {2}; Workman et al. (2017) [6] {15}
	Calypte anna	Anna's Hummingbird	Künstner et al. (2010) [4] {2}; Nabholz et al. (2011) [5] {2}
Casuariiformes	*Apteryx australis*	Southern Brown Kiwi	Subramanian et al. (2010) [5] {7}
	Apteryx owenii	Little Spotted Kiwi	Ramstad et al. (2016) [6] {27}
	Apteryx rowi	Okarito Kiwi[b]	Ramstad et al. (2016) [6] {27}
	Dromaius novaehollandiae	Emu	Künstner et al. (2010) [4] {2}; Nabholz et al. (2011) [5] {2}; Farlie et al. (2017) [6] {37}
Charadriiformes	*Gallinago media*	Great Snipe	Ekblom and Wang (2017) [8] {27}; Höglund et al. (2017) [6, 8] {27}
Columbiformes	*Columba livia*	Rock Dove[c]	Xu et al. (2016) [6] {18}; MacManes et al. (2017) [3] {2, 18, 23, 39}; Nimpf et al. (2017) [6] {45}
Falconiformes	*Falco cherrug*	Saker Falcon	Pan et al. (2017) [6] {27}
Galliformes	*Alectoris rufa*	Red-legged Partridge	Sevane et al. (2017) [1, 2] {3, 21, 22, 24}
	Coturnix japonica	Japanese Quail	Finseth and Harrison (2014) [1] {43}; Finseth and Harrison (2017) [6] {23, 43}
	Gallus gallus	Red Junglefowl	Bélteky et al. (2017) [6] {2}
	Meleagris gallopavo	Wild Turkey[d]	Harrison et al. (2015) [3] {18, 22, 23}; Wright et al. (2014) [3] {18, 22, 23}; Wright et al. (2015) [3] {18, 22, 23}
	Numida meleagris	Helmeted Guineafowl	Harrison et al. (2015) [3] {18, 22, 23}; Wright et al. (2015) [3] {18, 22, 23}
	Pavo cristatus	Indian Peafowl	Harrison et al. (2015) [3] {18, 22, 23}; Wright et al. (2015) [3] {18, 22, 23}
	Phasianus colchicus	Common Pheasant	Harrison et al. (2015) [3] {18, 22, 23}; Wright et al. (2015) [3] {18, 22, 23}
Passeriformes	*Catharus ustulatus*	Swainson's Thrush	Johnston et al. (2016) [7] {2}
	Chloris chloris[e]	European Greenfinch	Meitern et al. (2014) [2] {27}
	Corvus brachyrhynchos	American Crow	Künstner et al. (2010) [4] {2}; Nabholz et al. (2011) [5] {2}; Nabholz et al. (2010) [5] {2}
	Corvus corone	Carrion Crow[f]	Künstner et al. (2010) [4] {2}; Wolf et al. (2010) [5] {2}; Künstner et al. (2011) [4] {2}; Wolf and Bryk (2011) [3] {2}
	Corvus macrorhynchos	Large-billed Crow[g]	Vijayakumar et al. (2014) [1] {16}; Vijayakumar et al. (2015) [2] {16}
	Cyanistes caeruleus[h]	Eurasian Blue Tit	Künstner et al. (2010) [4] {2}; Nabholz et al. (2011) [5] {2}; Nabholz et al. (2010) [5] {2}; Künstner et al. (2011) [4] {2}; Mueller et al. (2015) [1, 3] {1, 2, 8, 10, 12, 13, 15, 16, 18, 19, 21–23}

Table 1 (continued)

Order	Species	Common name	Reference [research category] and {tissue}
	Eurillas virens[i]	Little Greenbul	Zhen et al. (2017) [1] {1, 2, 10, 15, 27}
	Ficedula albicollis	Collared Flycatcher	Ellegren et al. (2012) [1, 4] {2, 7, 13, 15–18, 21, 23}; Uebbing et al. (2013) [3] {2, 7, 13, 15–18, 21, 23}; Wang et al. (2015a) [5] {2, 7, 13, 15–18, 21, 23}; Uebbing et al. (2016) [5] {2, 7, 13, 15–18, 21, 23}; Wang et al. (2017a) [5, 6] {2, 13, 15–17, 21, 23}; Ålund née Podevin (2017) [6] {23}
	Ficedula hypoleuca	European Pied Flycatcher	Künstner et al. (2010) [4] {2}; Nabholz et al. (2010) [5] {2}; Künstner et al. (2011) [4] {2}; Nabholz et al. (2011) [5] {2}; Ellegren et al. (2012) [1, 4] {2, 7, 13, 15–18, 21, 23}; Wang et al. (2015a) [5] {2, 7, 13, 15–18, 21, 23}; Uebbing et al. (2016) [5] {2, 7, 13, 15–18, 21, 23}; Wang et al. (2017a) [5, 6] {2, 13, 15–17, 21, 23}; Ålund née Podevin (2017) [6] {23}
	Uraeginthus granatinus	Violet-eared Waxbill	Balakrishnan et al. (2013) [9] {2}
	Haemorhous mexicanus[j]	House Finch	Backström et al. (2013) [4, 6] {22}; Zhang et al. (2014b) [5, 2] {22}
	Hirundo rustica	Barn Swallow	Arai et al. (2017) [6] {40}
	Junco hyemalis	Dark-eyed Junco	Peterson et al. (2012) [1] {1, 2, 8, 10, 13, 15, 16, 18, 21, 23, 29–33}; Fudickar et al. (2016) [7] {17, 27}
	Manacus vitellinus	Golden-collared Manakin	Künstner et al. (2010) [4] {2}; Nabholz et al. (2011) [5] {2}; Nabholz et al. (2010) [5] {2}
	Melospiza melodia	Song Sparrow	Srivastava (2011) [5] {2, 10, 15, 17, 19, 34, 35}; Srivastava et al. (2012) [5, 8] {2, 10, 15, 17, 19, 34, 35}; Balakrishnan et al. (2014) [9] {6}
	Parus major	Great Tit	Santure et al. (2011) [1] {1, 2, 10, 13, 15, 18, 19, 21, 23}; Watson et al. (2017) [6] {15, 27}
	Passer domesticus	House Sparrow	Hagen et al. (2013) [8] {2, 10, 13, 15, 16, 23, 27}; Ekblom et al. (2014) [1] {3, 22, 27}
	Phylloscopus trochilus	Willow Warbler	Lundberg et al. (2013) [7] {2}
	Poecile atricapillus	Black-capped Chickadee	Cheviron and Swanson (2017) [6] {1}
	Serinus canaria	Atlantic Canary[k]	Lopes et al. (2016) [6] {15, 21}
	Setophaga caerulescens	Black-throated Blue Warbler	Kaiser et al. (2017) [5] {2, 15, 17, 27}
	Setophaga coronata	Myrtle Warbler[l]	Dick (2017) [7] {1}
	Sinosuthora webbiana[m]	Vinous-throated Parrotbill	Chu et al. (2012) [1] {2, 15}
	Spinus cucullatus	Red Siskin	Lopes et al. (2016) [6] {15, 21}
	Spinus spinus[n]	Eurasian Siskin	Videvall et al. (2015) [2] {27}
	Spinus tristis	American Goldfinch	Cheviron and Swanson (2017) [6] {1}
	Sturnus vulgaris	Common Starling[o]	Richardson et al. (2017) [1] {15}
	Taeniopygia guttata	Zebra Finch	Ekblom et al. (2010a) [1, 2, 6] {7, 15, 17, 21–23}; Ekblom et al. (2010b) [2, 4] {3, 7, 15, 17, 21–23, 27}; Gunaratne et al. (2011) [9] {2}; Balakrishnan et al. (2012) [3] {36}; Chen et al. (2016) [6] {21}; Chen et al. (2017a) [1, 6] {21}; Davidson and Balakrishnan (2016) [6] {2}; Newhouse et al. (2017) [2] {22}; Korlach et al. (2017) [1, 5] {2}; Han et al. (2017) [2] {3, 22}
	Turdus merula	Common Blackbird[p]	Franchini et al. (2017) [7] {27}; Koglin et al. (2017) [1] {1, 2, 10, 12, 13, 15, 16, 21–23, 27, 41, 42}
	Zonotrichia albicollis	White-throated Sparrow	Balakrishnan et al. (2014) [9] {6}
	Zonotrichia leucophrys	White-crowned Sparrow	Balakrishnan et al. (2014) [9] {6}

Table 1 (continued)

Order	Species	Common name	Reference [research category] and {tissue}
Psittaciformes	*Melopsittacus undulatus*	Budgerigar	Künstner et al. (2010) [4] {2}; Nabholz et al. (2011) [5] {2}; Cooke et al. (2017) [6] {40}
Sphenisciformes	*Pygoscelis papua*	Gentoo Penguin	Han et al. (2017) [2] {27}

Nomenclature according to the *IOC World Bird List* (version 7.3) (Gill and Donsker 2017). *Footnotes* indicate where nomenclature differs from the reference/references

Categorised research areas are given *in brackets*: *1* genome and transcriptome annotation; *2* immune response, disease biology and genetic disorders; *3* sex differences; *4* comparative genomics; *5* phylogenomics and population genomics; *6* trait evolution and evolutionary developmental biology; *7* migration; *8* genetic marker development; *9* bird song

Tissues used are given *in braces*: *1* breast muscle, *2* brain, *3* bursa, *4* caecal tonsil, *5* cerebellum, *6* duodenum, *7* embryo, *8* gizzard, *9* Harderian gland, *10* heart muscle, *11* ileum, *12* intestine, *13* kidney, *14* left optic lobe, *15* liver, *16* lung, *17* muscle, *18* ovary, *19* pancreas, *20* proventriculus, *21* skin, *22* spleen, *23* testes, *24* thymus, *25* thyroid, *26* trachea, *27* blood, *28* peripheral blood lymphocytes, *29* syrinx, *30* beak, *31* eye, *32* preen gland, *33* tongue, *34* egg, *35* bone, *36* cell line, *37* wing, *38* salivary gland, *39* pituitary gland, *40* feather, *41* stomach, *42* retina, *43* foam gland, *44* T-cells, *45* hair cells

[a]Chinese Goose

[b]Rowi

[c]White King Pigeon

[d]Turkey

[e]*Carduelis chloris*

[f]European Crow

[g]Jungle Crow

[h]*Parus caeruleus*

[i]*Andropadus virens*

[j]*Carpodacus mexicanus*

[k]Common Canary

[l]Yellow-rumped Warbler

[m]*Paradoxornis webbianus bulomachus*

[n]*Carduelis spinus*

[o]European Starling

[p]European Blackbird

Table 2 Reference transcriptomes available from the Ensemble database (www.ensembl.org) as of 18 December 2017

Common name	Latin name	Assembly	No. of annotated coding genes	No. of annotated non-coding genes	Tissues for which transcript analyses were performed	Evidence for genes/transcripts retrieved from
Chicken	*Gallus gallus*	Gallus_gallus-5.0	18,364	6492	1, 3–6, 8–11, 13–16, 18–22, 24–26	RNA-seq
Collared Flycatcher	*Ficedula albicollis*	FicAlb_1.4	15,303	6539	2, 7, 13, 15–18, 21, 23	RNA-seq
Mallard	*Anas platyrhynchos*	BGI_duck_1.0	15,634	567	2, 12, 16, 17, 22	RNA-seq
Turkey	*Meleagris gallopavo*	Turkey_2.01	14,123	755	NI	Online databases; protein, cDNA and expressed sequence tag sequences from turkey and chicken
Zebra Finch	*Taeniopygia guttata*	taeGut3.2.4	17,488	724	7, 15, 17, 21, 22,23,	RNA-seq

1 Breast muscle, *2* brain, *3* bursa, *4* caecal tonsil, *5* cerebellum, *6* duodenum, *7* embryo, *8* gizzard fat, *9* Harderian gland, *10* heart muscle, *11* ileum, *12* intestine, *13* kidney, *14* left optic lobe, *15* liver, *16* lung, *17* muscle, *18* ovary, *19* pancreas, *20* proventriculus, *21* skin, *22* spleen, *23* testes, *24* thymus, *25* thyroid, *26* trachea, *NI* no information, *RNA-seq* RNA-sequencing

Box 1. Characterisation of a transcriptome: a case study

Title: De novo assembly of the liver transcriptome of the European Starling, *Sturnus vulgaris*

Source: Richardson et al. (2017)

The European Starling *Sturnus vulgaris* is an important model for ecology and invasion biology (Richardson et al. 2017). To facilitate future molecular ecology and evolution studies in the European Starling, Richardson et al. (2017) characterised the liver transcriptome of this species. For this purpose, liver tissue was collected from two juvenile males from Western Australia. RNA was isolated from the tissues, and mRNA libraries were prepared and sequenced using the Illumina HiSeq 2500 sequencing platform. In total, 230 million paired-end sequencing reads of 125 base pairs (bp) were generated. The authors removed low-quality sequences and sequencing adaptors using the software Trimmomatic (Bolger et al. 2014), resulting in a total of 45 million high-quality reads for further analyses. The sequencing reads were assembled de novo using Trinity (Grabherr et al. 2011). After filtering out assembled contigs shorter than 300 bp, a total of 59,557 transcripts remained. Likely coding regions were identified using TransDecoder (https://github.com/TransDecoder/TransDecoder), and functional annotation of the transcriptome was undertaken using the Trinotate pipeline (http://trinotate.sourceforge.net). To estimate gene expression, the non-assembled transcripts were mapped to the contigs using Bowtie2 (Langmead and Salzberg 2012) and quantified using RNA-seq by expectation-maximization software (Li and Dewey 2011). The authors assessed the completeness and quality of the transcriptome using benchmarking universal single-copy orthologs software (Simão et al. 2015), which showed similar completeness values to two other passerine transcriptomes (Meitern et al. 2014; Richardson et al. 2017). Further, they compared their transcriptome to the European Starling genome and predicted transcriptomes using a BLAST (Altschul et al. 1990) approach for the assembled transcripts as well as a mapping approach (bbmap, https://sourceforge.net/projects/bbmap/) for the non-assembled sequencing reads. While the majority of the sequencing reads and contigs were found in the genome, a smaller proportion was found in the predicted transcriptome. The authors thus suggest that the predicted transcriptome is missing certain features, and highlight the importance of using RNA-seq data when annotating genomes (Richardson et al. 2017). The European Starling transcriptome is an important resource for understanding rapid evolution and adaptation in birds, and to improve the annotation of the European Starling genome. Ideally, further studies will include more tissues, as well as samples from female birds, to add data to the transcriptome dataset to identify a broader range of transcripts.

Functional studies (gene expression)

One of the main advantages of transcriptomic studies is that they allow for the quantification of gene expression, information which is absent from genomic studies. By comparing the gene expression of individuals living in different environments or at different life stages, the importance of differential gene regulation can be studied. In the following sections, we will give some examples of how comparative gene expression data have been used to study gene regulation in birds within fields such as evolutionary developmental biology, immunology, migration and population genetics.

Evolutionary developmental biology

Evolutionary developmental biology (Evo-Devo) aims at discovering how developmental processes evolved in organisms (Müller 2007). This involves comparisons of developmental processes of different taxa and the study of phenotypic change during development (ontogeny) and evolution. Phenotypic change is normally caused by two different types of mutations: mutations of the coding sequence, which can change the function of the gene itself; or mutations of regulatory regions, which can affect the expression of the genetic material (Necsulea and Kaessmann 2014). Comparative transcriptomics has become an important tool for Evo-Devo studies due to its potential for the study of the molecular basis of phenotypic evolution as well as the evolution of the transcriptome (Pantalacci and Sémon 2015). In this framework, transcriptomic data are compared between embryos or organs during their development in different species (Pantalacci and Sémon 2015), including birds (Chen et al. 2016; Xu et al. 2016). By comparing the transcriptome between pigeon ovaries pre- and post-ovulation, Xu et al. (2016) investigated what changes occurred in the gene expression profiles during different stages of ovulation. The authors found that immune genes were enriched in the transcriptome of the post-ovulation stroma. This suggests that the immune response is at least partly responsible for post-ovulatory follicle regression and elimination (Xu et al. 2016).

Disease studies

During the genomic era, extensive efforts have been invested into finding genetic variants associated with disease states. The majority of disease-associated single nucleotide polymorphisms (SNPs) detected in genome-wide association studies are located in non-coding regions of the genome (Freedman et al. 2011). This indicates that these genetic variants might influence gene expression levels rather than protein function (Costa et al. 2013). And indeed, by comparing gene expression profiles in healthy and diseased individuals we can get a better understanding of what parts of the genome are up- or downregulated during disease, or detect certain disease-causing alterations that might not be observed on a genomic level but can only be detected when looking at the expression of genes.

A commonly used experimental setup to study the immune response to infection is to undertake comparative gene expression studies in healthy and diseased individuals. This has been of particular interest in avian species that are known carriers of zoonotic diseases, as this gives an insight into how birds deal with these infections. A number of studies have used transcriptomics to study the immune response to infectious diseases such as avian influenza virus (AIV)

in non-model species of birds (Huang et al. 2013; Smith et al. 2015; Vijayakumar et al. 2015). A particular interest in understanding what makes a species resistant or susceptible to a certain infectious disease has further led to a number of comparative immunological studies between different species of birds. Using transcriptomics, Smith et al. (2015) found key differences in expression patterns of the interferon-induced transmembrane protein gene family in ducks and chickens, which the authors suggested might contribute to the large differences in susceptibility to AIV in these birds.

Another interesting application of transcriptomics, which has a great potential to advance the field of immunology as well as studies of host-parasite co-evolution, is dual RNA-seq of host and pathogen. Transcriptomic studies (and RNA-seq in particular) result in millions of sequence reads of which the vast majority will be from the target species sampled. A small fraction of the sequences, however, might represent transcripts from a previously known or unknown pathogen from the host tissue sampled. RNA-seq thus enables the simultaneous study of gene expression in the host and its pathogens, a possibility which was limited when using pre-NGS techniques due to the need for specific primers and hence prior knowledge of the pathogens present in the sample (Westermann et al. 2012). In a study on the effect of avian malaria infection on Eurasian Siskins *Carduelis spinus* the transcriptome was characterised from the host (Videvall et al. 2015) and the parasite (Videvall et al. 2017) simultaneously. Using this setup, the authors showed that a number of genes associated with functions within the immune system, stress response, cell death regulation, metabolism, and telomerase activity were overrepresented in the infected hosts (Videvall et al. 2015). The gene expression profiles of the parasites were more similar within the same individual across different infection stages than between individuals at the same infection stage, suggesting that the parasite might adjust its gene expression to specific host individuals (Videvall et al. 2017).

Migration strategies

A prominent aspect of avian biology that can be investigated using gene expression is bird migration. Dingle and Drake (2007) describe migration as an 'adaptation to resources that fluctuate spatiotemporally'. In birds, this could for example be the movement between nonbreeding areas and breeding grounds in response to seasonal change. Migratory birds go through behavioural, morphological and physiological changes before and during migration (Bowlin et al. 2010; Hedenström 2008). Using comparative gene expression, we can get a better understanding of how the genome is regulated at different stages of bird migration (Box 2). Johnston et al. (2016), using RNA-seq data, found 188 genes that were differentially expressed in the brain of captive Swainson's Thrushes *Catharus ustulatus* during different migratory states. A large number of these genes were associated with functions such as cell adhesion, proliferation and motility, and the authors hence suggested that migration-related changes might be regulated by seasonal neural plasticity (Johnston et al. 2016). However, migration is a complex behaviour, which certainly involves hundreds of genes, some of them with known functions (these can be identified by annotation with other known genes) and others, for which a function still needs to be explored.

Box 2. Gene expression during bird migration: a case study

Title: Animal tracking meets migration genomics: transcriptomic analysis of a partially migratory bird species

Source: Franchini et al. (2017)

The Common Blackbird *Turdus merula* shows a range of overwintering strategies, from migratory behaviour in Northern Europe to resident behaviour in Southern Europe. To investigate the genetic basis of these strategies, Franchini et al. (2017) used transcriptomics of blood cells to compare gene expression between resident and migratory Common Blackbirds in a partially migratory population of blackbirds in Southern Germany. Blood was collected from 12 individuals, and their migratory strategy was assessed using a combination of geolocators and radio transmitters. Briefly, 100uL blood was collected shortly before the winter migration, immediately frozen on dry ice, and subsequently stored at − 80 °C. RNA was isolated and mRNA libraries were prepared and sequenced on the Illumina HiSeq 2000 sequencing platform. A total of 59–101 million paired-end sequencing reads of 100 bp were obtained per individual. Low-quality reads and sequencing adaptors were removed using the CLC Genomics Workbench (CLC bio, Aarhus, Denmark). Further, overlapping paired-end reads were merged into a single read using SeqPrep (https://github.com/jstjohn/SeqPrep). To detect and remove potential contaminating sequences, the transcripts were searched against a custom-made database containing microbial sequences [source: National Center for Biotechnology Information (NCBI) reference sequence, September 2015]. The remaining 642 million sequencing reads were subsequently assembled de novo using Trinity (Grabherr et al. 2011). To identify genes, contigs of > 200 bp were blasted (Altschul et al. 1990) against protein sequences from chicken, duck, Zebra Finch, turkey and flycatcher, as well as mouse and human (Ensembl release 77). To quantify gene expression, the short sequencing reads from each individual were aligned to the transcriptome using Bowtie2 (Langmead and Salzberg 2012), and transcripts were clustered and subsequently quantified using Corset (Davidson and Oshlack 2014). Differential expression analyses were undertaken using the R package DESeq 2 (Love et al. 2014), and further validated using the R package EBSeq (Leng et al. 2013). The authors found differentially expressed genes related to hyperphagia, moulting and enhanced DNA replication and transcription when comparing resident and migrating individuals. These results contribute to our understanding of the molecular mechanisms involved in migratory behaviour in birds. One restraint of studies of this kind is that organs that might be of high relevance, such as brain, liver or intestine, cannot be investigated (for more details see "Tissues"). At the time of this study, there was also no reference genome or transcriptome available for the Common Blackbird. This gap in our knowledge, however, is currently being closed as a new Common Blackbird de novo transcriptome annotation has recently been published based on 14 different tissues (Koglin et al. 2017).

Population transcriptomics

Differences in gene expression have long been suspected to be highly important in evolutionary change, along with genetic divergence (Ellegren and Sheldon 2008; King and Wilson 1975). In population transcriptomics, gene expression patterns are compared between different individuals and populations, and can thus be used to study evolutionary processes such as adaptation and speciation. Indeed, large variations in gene expression have been detected between individuals and natural populations (Oleksiak et al. 2002; Whitehead and Crawford 2006). RNA-seq offers unique opportunities to compare gene expression between individuals and populations as well as species, partly because it does not depend on the development of primers with similar efficiency for the populations or species of interest. Further, RNA-seq allows for the detection of differences in, for example, isoform (see "Alternative splicing/transcriptome isoforms") and regulatory transcript (see "Coding and non-coding RNA") abundance. Recent developments in NGS technologies, as well as the decrease in sequencing costs, make RNA-seq a useful tool for population-wide comparisons because multiple individuals can now be subjected to sequencing for a price that was impossible to achieve just a few years ago ("Library prep" and "Sequencing depth and read length" sections). In order to study adaptation to high altitudes in the Saker Falcon, Pan et al. (2017) did a comparative transcriptomic study between populations inhabiting habitats of different elevations across Eurasia. At the nucleotide level, the authors found 37 SNPs in transcripts that were under directional selection in falcons that inhabited the high altitudes of the Qinghai–Tibetan Plateau, of which several were located in genes involved in oxygen transport and immunity. Further, the authors found that genes involved in oxygen transport were enriched and that half of the upregulated transcription factors were related to hypoxia responses. This nicely shows that changes both at the genomic and transcriptomic level have occurred in response to rapid adaptation to new environments. However, there are likely many other relevant genes, which could not be identified in such analyses, because corresponding genes might not have been analysed with functional assays in model organisms.

Gene interactions and function

One of the main advantages of RNA-seq is that up- or downregulation of all expressed genes can be detected simultaneously in a single RNA-seq run. This is of paramount importance, as a gene product in a pool of other translated genes usually acts as part of a complex network. For example, in order to fight an avian influenza virus infection, an immune response is induced through the upregulation of a network of genes; starting with the recognition of the pathogen and ending with an antiviral response. To find networks of genes and molecules that are over- (or under-) represented in a transcriptomic study, pathway and network analyses can be undertaken (Tomfohr et al. 2005). The *Kyoto Encyclopaedia of Genes and Genomes* is one example of a popular database resource that can be used to predict gene regulatory networks from the gene expression profiles (Kanehisa and Goto 2000; Kanehisa et al. 2010; Ogata et al. 1999) and has been so in birds (Lu et al. 2015; Peterson et al. 2012; Tariq et al. 2015; Vijayakumar et al. 2014). With the recent publication of the first goose genome (Swan Goose *Anser cygnoides*), Lu et al. (2015) used transcriptomic profiles to study the susceptibility of geese to fatty liver disease. Using pathway analyses the authors showed that the majority of genes in glucose and lipid metabolic pathways were upregulated in overfed individual geese (Lu et al. 2015). Although this study was undertaken in domesticated geese, understanding the mechanism behind weight gain in migrating species is of high relevance for ecology research, too.

Genetic marker discovery

While whole genome or transcriptome sequencing has greatly promoted our understanding of the biology of birds, many research questions of interest to ecologists are also answerable using only a subset of information from the genome or transcriptome. Genetic markers that represent information along the genome are routinely used to answer important research questions within population genetics and evolution (Davey et al. 2011; Morin et al. 2004; Schlötterer 2004; Selkoe and Toonen 2006). Most of the genetic markers used in bird studies derive from non-coding DNA sequences; however, the transcriptome can also be used for marker discovery. Sequencing the transcriptome instead of the entire genome makes the discovery of markers cheaper and thus more feasible for small research groups studying non-model species. Further, developing markers from a functionally relevant subset of the genome can be beneficial for studying adaptations (Wolf 2013). However, developing genetic markers from coding genes means that they are usually not

neutral, and this violates the assumptions of most population genetics models (Hartl and Clark 2007). Designing primers from exon regions alone might also cause problems when amplifying genomic DNA (gDNA), as primers for markers close to exon/intron boundaries might span long intron regions in the genome, which will hence not amplify successfully (De Wit et al. 2015).

In birds, RNA-seq data have been used to develop genetic markers such as microsatellites and SNPs. Microsatellites are regions in the genome composed of repeated mono-, di-, tri- or tetranucleoide motifs (Ellegren 2004). Microsatellites in repetitive DNA have a high mutation rate and are therefore highly polymorphic with different numbers of repeats in different individuals (Ellegren 2004), and are thus good markers for detecting genetic variation within or between closely related populations. This statement basically applies for non-coding DNA. However, RNA-seq data may be of interest in this context.

For House Sparrow *Passer domesticus*, a transcriptomic approach yielded 327 gene-linked microsatellites, thereby providing important genomic tools for future molecular ecology studies in this species (Ekblom et al. 2014).

SNPs are single base pair positions in gDNA where variation can be detected, and where the least common variant (allele, in this case called a 'minor allele') can be found in at least 1% of the population of interest (Brookes 1999). SNPs are abundant in the genome and can be found especially in non-coding but also in coding regions, and are hence valuable markers for many types of research questions or to see whether genetic variants are associated with a certain trait or disease. SNPs have been discovered using transcriptomics in many avian species (Balakrishnan et al. 2013; Ekblom and Wang 2017; Ekblom et al. 2014; Hagen et al. 2013; Kaiser et al. 2017; Lundberg et al. 2013; Ramstad et al. 2016; Santure et al. 2011; Srivastava 2011; Srivastava et al. 2012; Vijayakumar et al. 2014; Zhang et al. 2014b) (Box 3). Whole body transcriptomics was used to find candidate loci explaining the divergence between two subspecies of songbirds with several phenotypically different characteristics inhabiting either the mainland or islands of Alaska (Srivastava 2011). The authors found 1402 SNPs or indels that were fixed between populations and subspecies, and hence provided candidate loci for further evolutionary studies (Srivastava 2011).

Box 3. Genetic marker discovery: a case study

Title: Sixteen kiwi *Apteryx* spp. transcriptomes provide a wealth of genetic markers and insight into sex chromosome evolution in birds

Source: Ramstad et al. (2016)

The unique biology of kiwis *Apteryx* spp. make them interesting species for evolutionary studies. Further, all kiwis are currently considered 'vulnerable' or 'near-threatened' (*IUCN Red List of Threatened Species*, version 2017-3, 2017), making them important species for conservation biology. To develop genetic markers and hence enable population genomic, phylogenomic or molecular evolution studies in kiwis, Ramstad et al. (2016) sequenced transcriptomes of 16 kiwis from two species, the Little Spotted Kiwi *Apteryx owenii* and the Rowi *Apteryx rowi*. Note that the current *IOC World Bird List* (Gill and Donsker 2017) calls this species 'Okarito Kiwi'. We here use the vernacular name 'Rowi' as in the published paper. The authors collected 0.5 mL blood from eight Rowis and eight Little Spotted Kiwis from wild populations in New Zealand. RNA was isolated, and mRNA sequencing libraries were prepared and sequenced on the Illumina HiSeq 2000 sequencing platform. A total of 851,254,015 paired-end reads of 100 bp were obtained. To reduce the proportion of transcripts from haemoglobins in the assembly, transcripts mapping to haemoglobin sequences (Bowtie2, Langmead and Salzberg 2012) were removed and the remaining transcripts were assembled de novo for each species separately using Trinity (Grabherr et al. 2011). To assess the completeness of the transcriptomes, the Little Spotted Kiwi transcriptome was aligned using the North Island Brown Kiwi *Apteryx mantelli* genome (Le Duc et al. 2015) using Bowtie2 (Langmead and Salzberg 2012). To identify orthologous genes the assembled transcripts were searched against the NCBI protein database (Gish 1993), and the gene function of the ortholog hits was described using gene ontology analyses in Blast2Go (Conesa et al. 2005). To enable SNP detection the non-assembled sequencing reads were aligned to the Little Spotted Kiwi transcriptome using Bowtie2 (Langmead and Salzberg 2012). SNPs were then called using GATK (McKenna et al. 2010), and SNPs with a sequencing depth of $< 50\times$ in each individual and a distance of less than 200 bp from one another were removed. In total, the authors provide 120,035 SNP markers from the Rowi and 27,170 SNP markers from the Little Spotted Kiwi, of which roughly 67,000 SNPs were unique to the Rowi and hence can be used to differentiate the species (Ramstad et al. 2016). Out of these markers, 29,313 and 12,384 SNPs can be used to differentiate individuals within the same species for the Rowi and Little Spotted Kiwi, respectively (Ramstad et al. 2016). The higher number of polymorphic sites in the Rowi transcriptome in comparison to the Little Spotted Kiwi transcriptome is likely due to differences in demographic history between the species, and could be due to the fact that the Little Spotted Kiwi underwent a bottleneck about a century ago (Ramstad et al. 2016).

Genetic variation

Transcriptomic data represent a functional subset of the genome and offer unique opportunities to investigate genetic variation in coding and regulatory regions. This might be of particular interest when comparing phenotypes, which are likely to be caused by changes in the functional part of the genome. The functional part of the genome might further be under different evolutionary forces than the rest of the genome. Studying genetic variation in this part of the genome might hence provide a different angle to evolutionary studies than when using the whole genome. Transcriptomics have been used to study evolution, phylogenomics and population genomics in various avian species, as outlined below.

The study of speciation

In population genomics and phylogenomics, numerous loci or genome regions are studied simultaneously to disentangle questions about the importance of evolutionary processes such as mutation, genetic drift, gene flow and natural selection within and between closely related species (Luikart et al. 2003). The large amounts of data generated in whole-genome shotgun sequencing projects and the high costs of sequencing the entire genome make this approach problematic for population-wide studies in which tens of samples need to be analysed for each population. By sequencing a subset of the genome, the sequencing power can instead be focused on a small region of interest and more individuals can then be included for the same costs. As the transcriptome represents only the functional subset of the genome, the complexity is further reduced; thus there is great potential to study population genomics or phylogenomics using transcriptomic data (De Wit et al. 2012). One experimental difficulty is to obtain non-degraded RNA from wild birds (see "Preservation methods"). By comparing the genetic diversity of the transcriptomes of two differentially migrating subspecies of the Willow Warbler *Phylloscopus trochilus* Lundberg et al. (2013) confirmed results from previous studies that showed that the majority of the genetic variation is shared between the subspecies. The authors, however, also found a small set of SNPs that was differentiated between the subspecies. These SNPs clustered on two chromosome regions, and the authors suggest that these regions might be influenced by divergent selection associated with the subspecies' migration strategies (Lundberg et al. 2013). These results provide a starting point for further research to better understand the importance of these genome regions for bird migration. Similarly, transcriptomics was used for comparative genomic analyses between ten non-model avian species (Künstner et al. 2010), demonstrating its utility as a genomic resource for phylogenomics. In this study, the authors detected a negative correlation between chromosome size and the synonymous substitution rate, which suggests that small chromosomes have higher mutation rates than large chromosomes (Künstner et al. 2010).

Sex differences

Birds exhibit a wide range of sexual dimorphism, including differences in behaviour, physiology and morphology. Ornithologists have long asked how these differences arise, considering that the majority of the genome is shared between the two sexes.

At the genome level, sex differences are manifested in the presence of two structurally distinct sex chromosomes, which have arisen through different selection pressures between females and males. As in the XY sex-determination system of mammals, the avian sex chromosomes evolved from autosomal chromosomes and differentiated through stepwise suppression of recombination, which has led to the loss of most functional genes on one of the sex chromosomes (Ellegren 2011). In contrast to mammals, birds have evolved a female heterogametic sex system, with females having two distinct sex chromosomes (ZW) and males two copies of the Z chromosome (ZZ) (Ellegren 2000). The distinct pattern of inheritance in sex chromosomes means that they experience evolutionary forces different from the rest of the genome (Mank et al. 2010), and they have hence been subject to a number of evolutionary studies in birds (Wang et al. 2014; Zhou et al. 2014), also using transcriptomic data (Balakrishnan et al. 2013; Künstner et al. 2010; Ramstad et al. 2016; Wright et al. 2015). Wright et al. (2015) studied sequence and expression data of sex chromosomes simultaneously in six species of birds using RNA-seq data. The authors found that gene divergence between the species was higher on the Z-chromosome than on autosomal chromosomes. This is in line with the faster-Z evolution theory (Meisel and Connallon 2013).

A research question that has further received a lot of attention is how individuals of the homogametic sex avoid overexpression of the genes located on those chromosomes? In the XY sex determination system of mammals, one of the female's X chromosomes is silenced through a process called 'dosage compensation' into the Barr body (Charlesworth 1996; Ohno 1959). Interestingly, evidence of dosage compensation has not yet been found in the avian sex determination system (Baverstock et al. 1982). This finding has been supported in several species of birds such as the Carrion Crow *Corvus corone* (Wolf and Bryk 2011) and Collared Flycatcher (Uebbing et al. 2013), using transcriptomic data.

The lack or presence of certain sex chromosomes in an individual, however, cannot alone explain the vast differences seen between females and males. This is supported

by a number of studies that have shown sex-linked expression patterns in genes located on autosomes (as reviewed in Parsch and Ellegren 2013). Transcriptomic data enable genome-wide studies of sex-biased gene expression, and transcriptomics has therefore been an important tool in discovering what genes contribute to the different phenotypes seen in males and females (Mank 2009; Parsch and Ellegren 2013). Today, we still know little about how birds regulate the expression of the sex-linked genes, and we therefore expect to see more transcriptomic studies on this topic in the near future. In a comparative gene expression study in female and male Blue Tits *Cyanistes caeruleus*, 53% of all expressed genes showed sex-biased expression patterns (Mueller et al. 2015). Interestingly, a larger portion of the genes that were expressed at similar levels in both sexes mapped to annotated protein-coding genes than genes that were expressed in a sex-biased manner (Mueller et al. 2015). The authors hence suggested that sex-biased gene expression might involve non-coding and regulatory elements of the genome (Mueller et al. 2015).

Alternative splicing/transcriptome isoforms

In eukaryotes, protein-coding genes consist of exons which are used as templates to synthesise the polypeptide, and interspersing regions called 'introns' (Fig. 1). During transcription, pre-messenger RNA (pre-mRNA) molecules consisting of all exons and introns are transcribed. The pre-mRNAs then undergo a splicing event where introns are removed and the exons ligated, leaving as a result mature mRNA, which can be translated into proteins. In a process called 'alternative splicing', different combinations of exons can be ligated. Through this process several mature mRNA molecules (called 'isoforms') can be encoded by the same gene (Fig. 1). Alternative splicing is a common phenomenon in eukaryotes, and is believed to contribute to the complexity of the proteome in all species. In fact, about 40–60% of the human genes have differently spliced isoforms (Ast 2004). This explains the paradox between the high numbers of proteins in comparison to the moderate number of genes (e.g. 20,000 in vertebrates) (Modrek and Lee 2002).

RNA-seq has been an important tool for discovering alternatively spliced isoforms (Wang et al. 2009), especially in non-model species without a well-annotated reference genome. Zhang et al. (2014b) studied alternative splicing in spleen transcriptomes from two populations of House Finches *Haemorhous mexicanus* with different exposure histories to the bacterium *Mycoplasma gallisepticum*. The authors found a total of 41.8 and 40.8% transcripts, respectively, with two or more splice variants, of which 0.9% of the isoforms were found in one of the populations only (Zhang et al. 2014b). The high number of isoforms produced by the same gene, ranging from two to 23 variants, suggests that alternative splicing is likely also common in other avian species (Zhang et al. 2014b).

As most NGS technologies depend on the assembly of short sequences spanning the whole mRNA molecule, they are of limited use for accurately recognising different isoforms. In fact, only sequences covering the whole transcript can unambiguously recognise the true number of isoforms from a particular gene. New technologies enabling the sequencing of longer DNA fragments, such as IsoSeq (http://www.pacb.com/blog/intro-to-iso-seq-method-full-leng/) and MinION (https://nanoporetech.com/products/minion), will greatly improve the possibility of detecting different isoforms in transcriptomic studies. The application of these third-generation sequencing technologies to RNA sequencing are further discussed in "Choice of sequencing platform".

Planning an RNA-seq study: a quick guide

In this section we will briefly guide the reader through the most important steps involved in a RNA-seq workflow (Fig. 2), with the focus on non-model species research. We will further point the reader to more detailed literature for each step of the workflow. There are currently a number of reviews, tutorials and best practise guides available that should be used in parallel to find more in depth information on how to plan and smoothly undertake transcriptomics studies (Bullard et al. 2010; Conesa et al. 2016; Mazzoni and Kadarmideen 2016; Oshlack et al. 2010; Teng et al. 2016; Vijay et al. 2013; Wang et al. 2009; Wolf 2013). Eventually, early meetings with one's local sequencing core facility and/or a commercial provider should be sought for details.

Sample collection

Tissues

The content and composition of RNA is tissue specific, and hence the research question of interest will direct what tissue to sample for a particular study. For example, brain tissue might be the best tissue for studying bird song while intestine might be the optimal tissue for studying nutrition or the response to low pathogenic AIV. Many tissues, including brain and intestine, cannot be sampled from living individuals and this can become an issue when working with wild populations of birds and, in particular, when working with endangered species. If an individual has to be released after sampling, such as might be the case when working with protected species or when repeated measures are required, then tissues such as blood or feathers can be used for RNA isolation, although these transcriptomes might not be relevant for most research questions. When planning a transcriptomic

study under these conditions one hence needs to consider whether information obtained from the tissues used is relevant to answering the question of interest, and whether these tissues can be collected without harming the birds (McDonald and Griffith 2011). Further, the collection of samples from wild individuals usually requires special licences and the agreement of an ethical committee. Several studies have compared the transcriptomic profile between whole blood and other tissues during the immune response (Désert et al. 2016; Ekblom et al. 2014; Sandford et al. 2012), showing the utility of RNA from blood for disease ecology studies. Blood has further been successfully used to study sex chromosome evolution in two rare kiwi species, *Apteryx owenii* and *Apteryx rowi*, showing that this tissue can be used as a source of RNA for transcriptomic studies in protected birds (Ramstad et al. 2016).

Preservation methods

Each nucleotide in the RNA molecule contains a ribose sugar, which has a hydroxyl group that is prone to hydrolysis (Shukla 2015). This in combination with the rich abundance and high activity of ribonucleases, which actively degrade RNA (Ilinskaya and Mahmud 2014), make RNA less stable than DNA. Hence, specific care has to be taken when working with RNA to avoid degradation, both in the laboratory and in the field. In fact, changes in transcript abundance and complexity may occur during collection, handling and isolation of RNA (Lorkowski and Cullen 2006). To ensure that the analysed RNA accurately represents the in vivo expression profile of the sample, it is essential to stabilise the RNA sample as soon as possible after collection (Lorkowski and Cullen 2006). The most reliable method to avoid degradation of RNA after sampling is rapid freezing in liquid nitrogen (Wolf 2013). Another option, which might work under special conditions, is the use of RNA-stabilising reagents. Several commercial RNA-stabilising reagents are currently available, of which one of the most commonly used is RNAlater (Ambion). This stabilises and protects the integrity of RNA at room temperature for up to 1 day at 37 °C, 1 week at 18–25 °C or several weeks at 4 °C (Ambion). Cheviron et al. (2011) examined how preservation method and time between sample collection and processing affect the RNA quality and yield in avian tissue samples collected in the field. The authors found that RNA yields were higher for tissues that had been snap-frozen; however, the yield and quality of the RNA from tissues collected in the common standard chemical RNAlater and stored for up to 2 h before shock-freezing were also sufficient for most gene expression applications (Cheviron et al. 2011). In order to obtain good results from RNA-seq, the RNA must be of very good quality; thus it is important to optimise the preservation step in any transcriptome study.

RNA isolation and library preparation

Coding and non-coding RNA

Many of the RNA-seq studies have focused on mRNAs used for protein synthesis. However, RNA-seq can also provide information on RNA referred to as 'non-coding', i.e. RNA that will not be translated into proteins. Non-coding RNAs such as long non-coding RNAs (> 200 nt), small nuclear RNAs (snRNAs < 200 nt), microRNAs (miRNAs ≈ 22 nt), and small interfering RNAs (siRNA ≈ 21 nt) have important roles in biological processes such as gene regulation, replication, mRNA processing, and splicing, and might hence be of interest for certain research questions (Clancy 2008; Morris and Mattick 2014). When designing an RNA-seq study, it is important to keep in mind what type of RNA can best be used to answer the question of interest, as the RNA isolation step, the library preparation and the sequencing settings all depend on the type of RNA to be studied. The most common RNA species to be targeted in RNA-seq studies are mRNA, total RNA or small RNA (sRNA; including the sRNA types such as snRNA, miRNA and siRNA).

There are several aspects to take into account when choosing the population of RNA for a specific study (Kratz and Carninci 2014). mRNAs carry the genetic information from protein-coding genes in the genome and are used as a template for proteins during translation. The presence of certain mRNAs hence indicates what protein-coding genes are being expressed in a sample. As the goal of many transcriptomic studies is to see which genes are down- or upregulated under certain conditions, the use of mRNAs is a common choice. Further, when working with non-model species, annotation of other populations of RNA might be difficult and hence complicate further the interpretation of the data. However, mRNA is only a small portion (1–5%) of all RNA in the cell (Rao et al. 2006), and when overlooking other types of RNA the complexity of the entire RNA composition is underestimated. In order to capture a wider range of RNAs total RNA can be targeted, which allows for the detection of coding and multiple forms of noncoding RNA. This has been done in some avian RNA-seq studies (Chen et al. 2017a; Srivastava 2011; Videvall et al. 2015), which has contributed to our understanding of the role of functional RNAs (fRNAs) in birds. While whole-transcriptome analysis with total RNA gives a more comprehensive picture of the transcriptome than mRNA sequencing, certain types of RNAs will not be accurately represented in a total RNA sample. In particular, sRNAs such as miRNAs, which are about 22 nt in length (Bartel 2004), are easily lost during RNA isolation, and several companies have thus released isolation kits that are specific to the capture of sRNAs. In birds, upregulation of miRNAs has been observed in chicken and ducks infected with AIV (Li et al. 2015; Wang et al. 2012),

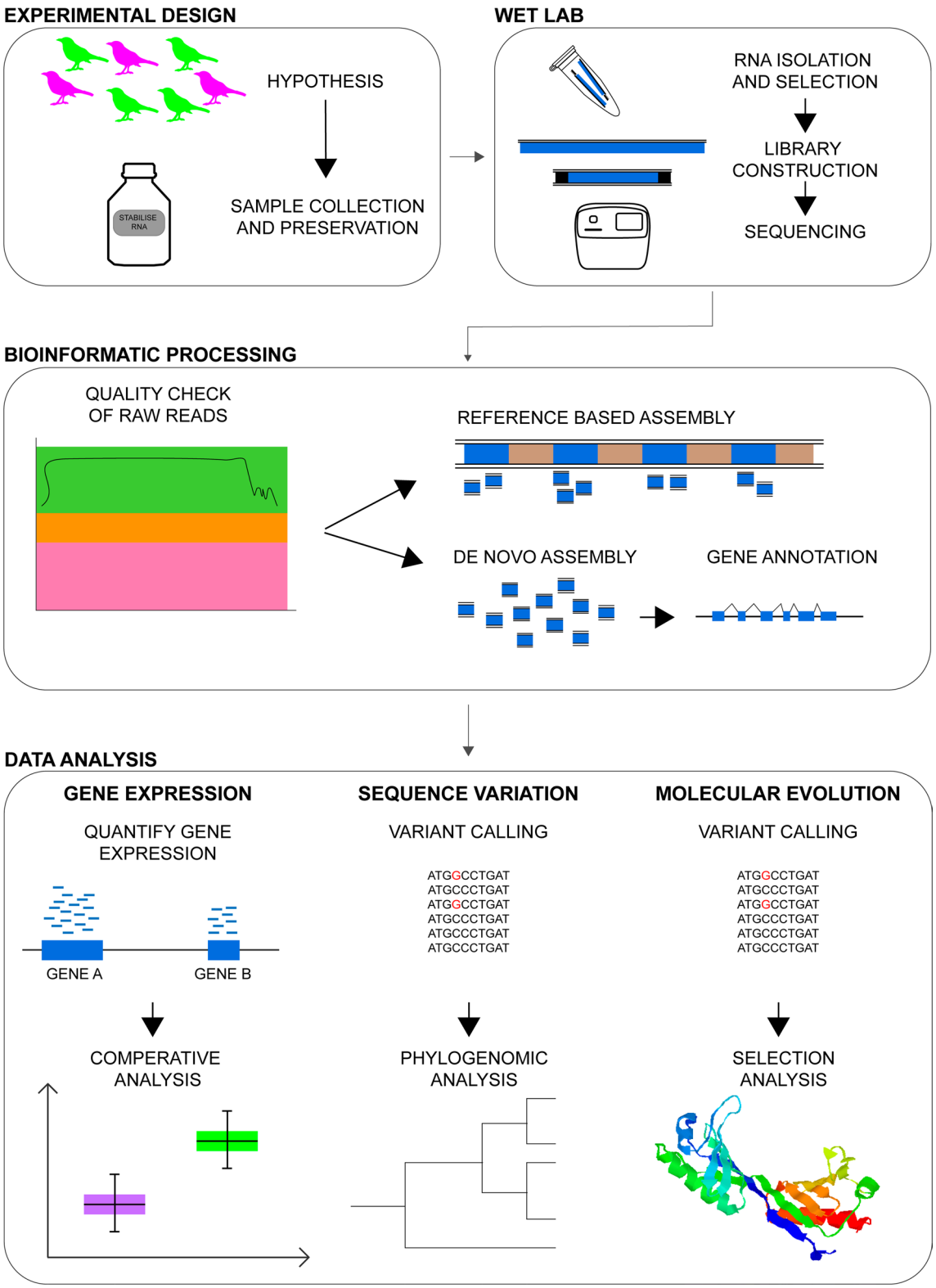

Fig. 2 Overview of a typical RNA sequencing workflow. The individual elements of the workflow are detailed in "Planning an RNA-seq study: a quick guide". Protein structure by Richard Wheeler (Zephyris), licensed under Creative Commons 3.0, Wikimedia Commons

and in Eurasian Siskins infected with avian malaria (Videvall et al. 2015). Further, the miRNA profile was altered in the auditory forebrain of Zebra Finches when they heard bird song (Gunaratne et al. 2011). This shows that miRNAs are involved in various biological functions in birds, ranging from the immune response to song communication, and that they are likely involved in more functions as well, although they have largely been overlooked so far.

RNA isolation

There is a wide range of RNA isolation kits available on the market, some of which are specific for a specific type of tissue and others specific for capturing RNA molecules of certain sizes (see Sect. "Coding and non-coding RNA"). It is important to keep in mind that most commercial kits have been developed for model species such as mouse or human, and that the protocols might in certain cases require optimisation to work optimally on bird tissues. One such example is blood, where avian red blood cells are nucleated in contrast to mammalian red blood cells and hence can block the columns that are used in most RNA isolation kits. In several avian RNA-seq studies using whole blood, a combination of kits was used to avoid this issue (i.e. Franchini et al. 2017; Meitern et al. 2014; Videvall et al. 2015). To minimise the risk of sequencing a mixture of unknown parts of gDNA and cDNA, gDNA should be removed during RNA isolation (Dotti and Bonin 2011). This can, for example, be achieved by the digestion of DNA with the enzyme DNase during RNA isolation.

Before library preparation the purity and degradation of the samples should be assessed using an automated electrophoresis system. Further, as many RNA-seq applications are dependent on quantification measures it is extremely important that the same amount of input RNA is used for the library preparation. The concentration of the samples should hence be carefully measured, preferably using sensitive fluorescence-based techniques.

Library preparation

During library preparation, RNA is converted to cDNA with a number of molecular adaptors that enable sequencing of the RNA of interest on the sequencer of choice. When sequencing mRNA or total RNA a typical library preparation protocol will start by either capturing mRNA molecules or removing rRNA (Fig. 3). In the case of mRNA selection, magnetic beads containing oligodeoxythymidylate molecules are usually used to capture RNA molecules with a polyadenylated region, which is present in most mRNA. In the case of total RNA sequencing, rRNA, which comprises about 90% of the RNA in the cells (Wilhelm and Landry 2009), is removed. In both applications, the RNA is thereafter fragmented, reverse transcribed to double-stranded cDNA, and adaptors allowing for sequencing are added to the cDNA. To enable pooling of samples during sequencing, sample-specific adaptors (also called 'indexes' or 'barcodes') can further be added to the samples. cDNA molecules with the appropriate adaptors are then enriched using PCR. Finally the quality of the libraries is validated using an automated electrophoresis system, normalised and, in many cases, pooled prior to sequencing. When interested in miRNA the RNA is not fragmented in the beginning of the protocol, but a size-selection step should instead be undertaken using a gel electrophoresis step (Fig. 3).

Many total RNA isolation kits provide information on strand orientation of the transcripts. This has several advantages, as it allows for identification of antisense transcripts, which might have regulatory roles, and because it might help resolve the difficulty of knowing what gene a certain sequence belongs to in the case of overlapping transcripts (Levin et al. 2010).

Sequencing facilities normally offer library preparation as a service, a good alternative for groups that are not interested in or do not have the appropriate facilities or experience for library preparation.

Sequencing

Sequencing depth and read length

In most next-generation sequencing projects, the genetic material is fragmented into short pieces, which are then sequenced in a random order. The sequencing platform used in a project sets the limit on the total number of fragments that can be sequenced in one sequencing run. By choosing different sequencing platforms and by pooling different numbers of samples in one sequencing run the sequencing depth per sample can thus be adjusted. The sequencing platforms also differ in their ability to sequence DNA fragments of different lengths. Depending on the project in mind one hence has to decide what sequencing length to aim for.

An ideal transcriptomic study would contain many experimental replicates with high sequencing depth of long sequences. As will soon become clear, when planning an RNA-seq study, however, sequencing is still costly despite recent advances. A trade-off between numbers of experimental replicates, sequencing depth and lengths, therefore, has to be made. If the goal of the study is to make a reference transcriptome then aiming for longer sequences might be the appropriate choice, as this will facilitate the de novo assembly. For this reason, many avian studies have used the now deprecated Roche 454 sequencing technology where sequence lengths can be several hundreds of nucleotides long (Ekblom et al. 2014; Peterson et al. 2012). However, if a good reference genome or transcriptome is already

available for the target species or a closely related one then shorter sequences (as generated by Illumina sequencers) might be sufficient—and much cheaper. The sequencing depth necessary will also differ from study to study. If some transcripts are known or expected to occur in low abundance compared to all the other expressed RNA in a cell then a high sequencing depth might be required, as the chance of sequencing rare transcripts will increase with the number of sequences. However, if the main goal is to find the most differentially expressed genes between two treatments and rare transcripts are of little interest, then fewer sequences might be sufficient. Around 100 million sequences of 100 nt have been shown to efficiently capture the majority of the genes in a sample, which is of importance when characterising the transcriptome (Vijay et al. 2013). In chicken, it was shown that 10 million sequencing reads of 75 nt captured about 80% of the annotated chicken genome, and the authors hence suggested that RNA-seq at this depth is a good alternative to microarray technology (Wang et al. 2011). The required sequencing depth will further determine how many replicates can be pooled in one sequencing run. Many sequencers have separate sequencing lanes, which can be used to increase the output or separate samples. Further, most companies offer index adaptors (so-called barcodes) that allow for pooling of up to 96 different samples (or, if custom made, even more), which can reduce the sequencing cost greatly. There are a number of tools that can be used to help estimate the number of replicates and sequencing depth necessary for a study design (i.e. Busby et al. 2013; Liu et al. 2013).

Choice of sequencing platform

In this review, we will only briefly discuss the available sequencing platforms and the possibilities and challenges that these platforms present for non-model research, as there are excellent reviews available that give deep insight into these technologies (Chu and Corey 2012; Metzker 2010; Morozova et al. 2009; Wang et al. 2009; Wolf 2013).

The most commonly used sequencing platforms for avian transcriptome research are Roche 454/pyrosequencing and Illumina (Table 3). The Roche 454/pyrosequencing technology was the first commercial NGS technology used for whole-transcriptome shotgun sequencing. This technology allowed for the sequencing of relatively long fragments, which is very useful for the de novo assembly of transcriptomes. The Illumina technology specialises in producing a high number of short sequencing reads. This has been of great value for reference-based RNA-sequencing projects, where the main purpose is to study or compare gene expression between treatment groups. Illumina is currently the most commonly used sequencing platform for RNA-seq studies, and there are a number of Illumina sequencing platforms available such as the MiSeq and the HiSeq series. The Illumina sequencing systems all have different specifications and flexibility, which will make them more or less suitable for a particular study designs. The MiSeq has short sequencing run times (4–56 h) and can sequence fragments of lengths 36–300 bp (single- or paired-end reads), and depending on the settings produce an output of about 540 Mb–15 Gb. The HiSeq series, which contains a number of sequencing systems (currently HiSeq 2500/3000/4000),

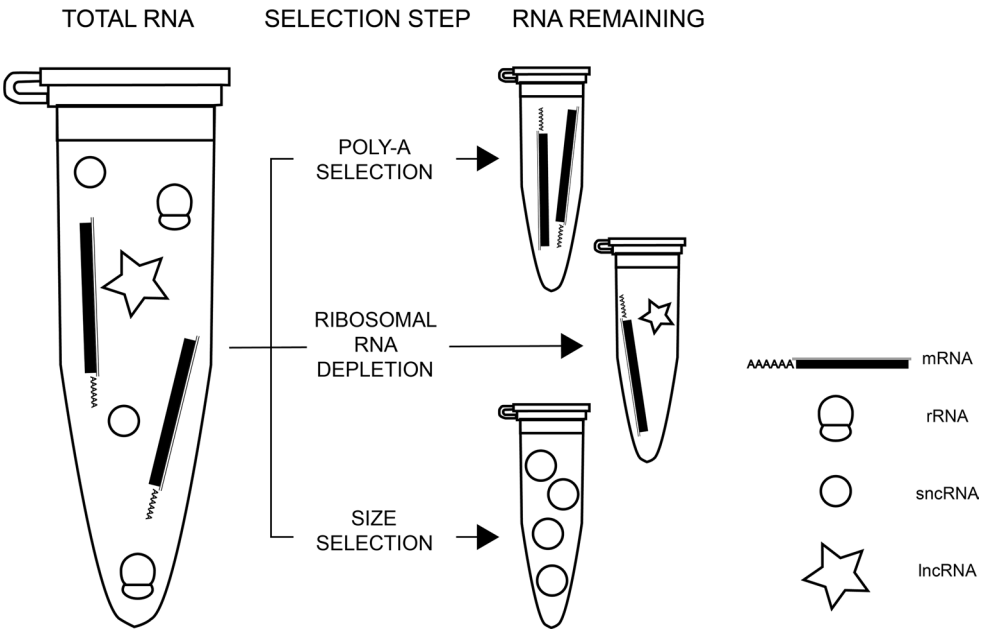

Fig. 3 The most commonly studied types of RNA in RNA-sequencing projects and how they are selected for. *mRNA* Messenger RNA, *rRNA* ribosomal RNA, *sncRNA* small non-coding RNA, *lncRNA* long non-coding RNA

offers great flexibility in sequencing lengths (single end 36 bp-paired end 250 bp) as well as total output (from nine to 1500 Gb), with sequencing run times ranging from 1 to 6 days. The continuous development of sequencing systems as well as reagent kits leads to rapid changes in sequencing outputs as well as costs; we therefore encourage those planning an RNA-seq study to contact their local sequencing facility to find the most suitable and cost-efficient option available for the study design.

While the platforms mentioned above differ in some aspects, they also have commonalities. For example, they use a fragmented cDNA library, which requires that the full-length RNA molecules are then reconstructed from the short sequences by further processing by software that can find overlaps between the individual sequences and stitch them together, a process called 'sequence assembly'. New technologies such as the Pacific Biosciences (PacBio) Isoform Sequencing (Iso-Seq) application (Gonzalez-Garay 2016), can generate full-length cDNA sequences, and hence reduce the need for reconstructing the full-length RNA molecule by post hoc assembly approaches (Rhoads and Au 2015). The first avian transcriptomes have recently been sequenced using this new technology (Kuo et al. 2017; Workman et al. 2017). Another new application, which enables sequencing of full-length transcripts, is the MinION from Oxford Nanopore Technologies (Ayub et al. 2013; Garalde et al. 2016). However, to our knowledge this application has not yet been used for RNA sequencing in bird species.

Transcriptome assembly or alignment

The most commonly used RNA-seq technologies produce a large library of short sequences, which have to be merged in order to reconstruct the original sequence (Oshlack et al. 2010). To ensure the best alignment possible, the quality of the sequences can be checked in a process called 'preprocessing'. During this process, quality scores which have been assigned by the sequencer can be used to evaluate the quality of the sequencing library, and sequencing adaptors which have not been successfully removed can be detected. Using this information, thresholds can be set to trim the sequences accordingly, thereby minimising the use of low-quality sequences.

There is a vast array of tools available for alignment and assembly of the sequences (Bao et al. 2011; Trapnell and Salzberg 2009), which will be more or less suitable for the project in mind. If a reference genome or transcriptome is not available then the sequences will have to be assembled de novo, which requires special de novo assembly software (Haas and Zody 2010). The access to a reference genome will enable alignment of the reads to the reference (Vijay et al. 2013). Alternatively, the sequences can be aligned to closely related species with a reference genome. Alignment tools differ in their ability to align sequences to references of more or less divergent species, and this will thus also affect what tool to use (Wolf 2013). In general, the quality of transcriptomes reconstructed with the help of reference genomes or transcriptomes is better than that of those reconstructed de novo. Once the sequences have been successfully aligned or assembled, genetic markers can be developed or transcript levels can be estimated.

Analysis

Variant calling

If transcriptomic data are available from several individuals then sites that are polymorphic within and between individuals can be detected from the aligned sequences. As misaligned sequencing reads could cause an overestimation of polymorphic sites it is important to have an alignment of high quality when calling SNPs. There are a number of tools available to detect SNPs using NGS data (reviewed in Nielsen et al. 2011). Using these SNPs the genotypes can be estimated for all individuals, and population genomic, phylogenomic or molecular evolution analyses can be undertaken. There are several reviews and tutorials that go into depth on how to call SNPs from NGS data, and how to use this to study population genomics and speciation in non-model species (De Wit et al. 2012, 2015; Toews et al. 2015).

Gene expression quantification

To quantify gene expression levels the number of sequencing reads that map to genes along the genome are counted. As the raw counts will vary with the sequencing depth of the samples, the raw counts have to be normalised to enable comparison across samples of interest. This is done by calculating the ratio of expression for each gene, taking the total number of sequencing reads per sample into account. Some of the most commonly used scales for gene expression are reads per kilobase of transcript per million mapped reads, fragments per kilobase of transcript per million mapped reads, transcripts per million and counts per million. The length of the gene can also be taken into account, as longer genes with a certain expression level will have more sequencing reads mapping to them than shorter genes with a similar expression level. Further, as highly expressed genes will affect the ratio of the remaining genes, the distribution of expression levels for each sample should also be normalised. This can be done using scale normalisation factors such as the trimmed mean of M values (Robinson and Oshlack 2010), the upper quartile (Bullard et al. 2010), or the geometric mean (Anders and Huber 2010). Once gene expression levels have been quantified and appropriately normalised for all samples, comparative analyses can be

Table 3 The most commonly used sequencing platforms in transcriptomics

Sequencing platform	Pros for non-model research	Cons for non-model research
Illumina	High number of short sequences make this technology attractive for comparative gene expression studies in species with a reference genome. Further, it is currently the cheapest technology per nucleotide (nt) sequenced	The short read lengths make this technology problematic for species without a reference genome/transcriptome
454/Pyrosequencing	The relatively long reads (~ 700 nt) facilitated de novo assembly, and this technology was hence suitable for species without a reference genome	This technology has been discontinued https://www.genomeweb.com/sequencing/roche-shutting-down-454-sequencing-business
PacBio	The possibility to sequence whole transcripts (several kilobases in length) makes this technology highly suitable for de novo assembly or for detecting alternative splicing	This technology is still new and costly

undertaken to estimate differential gene expression under the conditions of interest. Many types of software are available for this purpose (for more details see Soneson and Delorenzi 2013; Teng et al. 2016).

During differential expression analysis, the expression values for each gene are compared across the different conditions and will hence result in a significance value (often a *p*-value) for the comparison for each gene. As explained by Noble (2009), *p*-values are, however, only statistically valid when a single test is undertaken. Due to the large number of genes for which the expression levels are estimated and compared in RNA-seq projects, multiple testing corrections need to be applied. Several reviews go into more detail on how to analyse RNA-seq data, and we thus recommend further reading for a more in depth description of this complex subject (Haas and Zody 2010; Ockendon et al. 2016; Oshlack et al. 2010; Vijay et al. 2013).

Biological interpretation of RNA-seq data

To make sense of the data received from a transcriptomic study it is necessary to identify which genes the individual transcripts belong to and infer their function. When aligning sequences to well-annotated reference genomes, information on the known regions from the reference will automatically be assigned to the aligned reads, which greatly facilitates biological interpretation. In the case of de novo assemblies, an extra annotation step is necessary. During annotation, protein and transcript information from the target species or (closely) related species are aligned to the genome, and this information in combination with gene prediction is used to identify genes and to assign biological information to the gene lists (Yandell and Ence 2012). Several annotation pipelines that make use of a number of tools are currently available (Andersson et al. 2015; Cantarel et al. 2007). Still, gene annotation is anything but trivial, and hence having access to a well-annotated reference genome is highly beneficial. Only a fraction of the published avian genomes are well annotated (Table 2) and there is thus a great need for further work on many genomes. Further transcriptomic studies play a major role here. It should not be overlooked that even in well-studied model organisms (including humans) the function of many genes is still obscure. Unfortunately, even more genes need to be characterised in terms of function in birds.

Gene ontology (GO) analyses can be used to interpret the biological function of the list of genes retrieved from the alignment or annotation step. During GO analysis, genes are classified into ontology classes that describe gene products in terms of their associated molecular function, cellular component and biological process (Gene Ontology Consortium 2004). This classification scheme was initiated to generate consistent descriptions of gene products that can be used across species boundaries (Gene Ontology Consortium 2004). GO analyses are frequently used in transcriptome studies to help reduce complexity and highlight biological processes (Young et al. 2010), and also, of course, in avian studies (Balakrishnan et al. 2013; Chu et al. 2012; Ekblom et al. 2014; Peterson et al. 2012; Santure et al. 2011; Vijayakumar et al. 2014; Wright et al. 2015). By classifying the genes that were differentially expressed between migratory and sedentary Dark-eyed Juncos *Junco hyemalis* through GO analysis, Fudickar et al. (2016) found that genes involved in lipid transport and metabolism were overrepresented in migrant individuals while genes involved in reproductive processes were overrepresented in resident individuals.

Challenges of RNA-seq and its utility for non-model ornithology

NGS has revolutionised the study of transcriptomics. However, there are significant challenges that need to be taken into consideration when using RNA-seq technology for research on non-model organisms.

Library preparation: possible bias in the data

Some of these challenges are related to complications associated with the construction of libraries for the major NGS platforms. For example, the current fragmentation methodologies can introduce bias during template fragmentation (as discussed by Wang et al. 2009). Furthermore, a number of errors, such as self-priming, template switching, and reverse transcriptase inaccuracies can occur during cDNA synthesis (Ozsolak and Milos 2011; Wang et al. 2009). Also, most current RNA-seq protocols contain a template-amplification step, and as certain transcripts might be more prone to amplification than others, this step can lead to the uneven distribution of transcripts which is not representative of the sample (Garalde et al. 2016; Kozarewa et al. 2009). New technologies are continuously being developed to avoid or minimise these issues, including long-read technologies (Gonzalez-Garay 2016), direct RNA sequencing technologies (Ayub et al. 2013; Garalde et al. 2016; Ozsolak et al. 2009), and PCR-free protocols (Mamanova et al. 2010). Ensuring a high -quality library is of highest priority in all NGS projects, and it is therefore likely that we will see further developments aimed at minimising the problems involved in library preparation for RNA-seq within the near future. It is further important to stress that NGS needs DNA or RNA of high quality and purity.

Working with large datasets: big data, a new era for biology

Once a transcriptome has been successfully sequenced, the next challenge is to analyse the great amount of data received from deep-sequencing projects (for a detailed tutorial see Wolf 2013). A common RNA-seq project can easily result in a raw dataset in the order of 10–100 billion bases (Gb), and during analysis, this further expands due to intermediate results in the down-stream analysis pipelines. It is therefore important to provide for sufficient computational storage, either locally or via a remote high-performance computer cluster (Schatz et al. 2010). While some tools, such as Galaxy (https://usegalaxy.org/), offer a user-friendly web interface for analysing RNA-seq data (Goecks et al. 2010), most applications used during the analysis of RNA-seq are either run on the command line in the UNIX operating systems or in the R environment (R Core Team 2014) and thus require some bioinformatics expertise. Recent improvements in sequencing technologies have led to rapid advances in the availability of bioinformatics tools (Conesa et al. 2016; Schurch et al. 2016; Seyednasrollah et al. 2015), as well as user-friendly manuals (Law et al. 2016; Pertea et al. 2016) and tutorials (Conesa et al. 2016; Wolf 2013). Finally, it is imperative to make oneself familiar with bioinformatics or seek collaborations when designing a project.

Transcriptomics: from the laboratory to the field

The majority of the transcriptomic studies in birds have been performed under controlled laboratory conditions, and for a good reason. In these setups, one typically compares an experimental condition with a non-manipulated control, and by keeping all factors except the one of interest constant across the different treatment groups the influence of other factors is kept to a minimum. For example, by performing experiments between the treatment and control concurrently, or at the same time of day, temperature or diurnal rhythm can be controlled for.

When undertaking experiments in a controlled environment, the main idea is to reduce the complexity of the environment to make it possible for one to focus on one factor alone. However, wild birds live in a complex world, where their gene expression is influenced by a number of factors simultaneously. In order to better understand the flexibility of the transcriptome in birds, it is necessary to study transcriptomics under the wide range of conditions that they experience in their natural environment. In natural populations, it is, however, difficult to know for sure whether the differences observed between individuals or treatment groups are indeed due to the factor of interest or other factors that can neither be measured nor controlled for. Larger sample sizes will be needed in natural settings. It seems that a combination of controlled and natural experiments will give the best opportunities to better understand how the genome is regulated in different contexts.

Experimental proof as a last step to ascribing biological function

Knowing that a gene is up- or downregulated under a certain condition does not necessarily verify the function of that gene. Instead, the results can be used to propose a hypothesis about the function of a specific gene in a particular context. Experimental testing of such hypotheses is the actual proof beyond correlation from gene expression studies. Modification of specific genes using gene-editing technologies is currently the gold standard for determining gene function (Capecchi 2005). By this strategy, so-called knock-out or knock-in strains are produced, which are then available for biological and pharmacological research. These methods are labour extensive and delicate, also from an ethical perspective, and have been available only to dedicated research groups (Kratochwil and Meyer 2015). They are also particularly challenging in birds because of the inaccessibility of the zygote, which is enclosed in the egg (Cooper et al. 2017).

Recently it was shown that genome editing using homologous recombination can be undertaken in cultured chicken primordial germ cells (PGCs), which can then be injected into a surrogate egg shell (Schusser et al. 2013). Similar to

in mammals, breeding schemes can then be used as a measure to produce homozygous knock-out birds. This method was used to study the knock-out effect of an immunoglobulin which has an important role in the adaptive immune response (Schusser et al. 2013), and for the knock-out of the egg white protein ovalbumin (Park et al. 2014) in chicken. Later, sperm was successfully used as a delivery mechanism for gene-editing vectors in chicken, something which opens up the possibility for gene editing in species of birds where methodologies for the long-term culture of PGCs do not exists (Cooper et al. 2017). New technologies such as CRISPR–Cas9 for genome engineering have further made genome editing more feasible for research in non-model species (Doudna and Charpentier 2014), and have recently been used in chicken as well (Abu-Bonsrah et al. 2016; Oishi et al. 2016). The availability of a knock-out system in chicken has the potential to improve our general understanding of gene functions in birds, and allows for the exploration of whether this technique can be used in other species of birds as well.

Future

Single-cell sequencing

Cells are the building blocks of multicellular organisms and usually highly specialised (more than 200 cell types occur in vertebrates). In most transcriptome studies, the gene expression profile has been studied in tissue samples comprising a wide range of different cells and cell populations. The observed transcriptomes are thus a result of a cocktail of transcripts from a large number of cells of different kinds, and specific information on gene expression in certain types or subpopulations of cells is lost. In the past few years, several technologies enabling gene expression profiling in single cells have been developed (Hashimshony et al. 2012; Jaitin et al. 2014; Kolodziejczyk et al. 2015; Picelli et al. 2013; Tang et al. 2009), which is a major breakthrough for biology. These technologies allow for a more in-depth knowledge of basic research questions such as how many distinct cell types are there and how do they interact with each other, as well as more specific research questions such as why are some cells affected by diseases when others are not (Eberwine et al. 2014)? The majority of these technologies require cell sorting prior to use, but this is difficult in non-model species for which cell-sorting protocols have not been developed. New technologies allowing for single-cell gene expression analysis in thousands of unknown types of cells, however, have now also made single-cell gene expression analysis possible in non-model species (Macosko et al. 2015). We believe that we will see more single-cell RNA-seq studies in birds within the near future.

Spatially resolved *omics

As mentioned previously, the majority of transcriptomic studies are undertaken on homogenised tissues. This does not only lead to a loss of cell-specific information but also the loss of positional information of the expression pattern in the tissue of interest. Maintaining spatial information on gene expression of single cells or subpopulations of cells in tissues could help us to better understand how different cells function and are regulated, where they are localised and how they interact in complex tissues (Crosetto et al. 2015). Some techniques can be used in combination with RNA-seq such as laser-capture microdissection, where single cells or subpopulations of cells can be harvested from tissue samples and used for downstream analysis (Espina et al. 2006); microtomy sequencing, where RNA is extracted from thin cryosections (Junker et al. 2014); or spatial transcriptomics where tissues are positioned on an array with spatially barcoded primers, which allow for two-dimensional positional information to be taken into account in the analysis (Ståhl et al. 2016). These technologies offer new possibilities to learn more about avian biology, in particular within areas such as neurobiology and immunology.

On-site sequencing platforms

One of the newly available sequencing instruments which has great potential for future RNA studies in a field setting is the portable sequencing device MinION (Oxford Nanopore Technologies, Oxford, UK). The current version of the MinION weighs less than 100 g and is powered by the USB port of a laptop, thus can easily be taken into the field and used on-site. Having the possibility of sequencing the sample of interest on-site does not only have the potential to reduce the time between harvesting and sequencing, but can also facilitate the workflow in countries where sequencing facilities are scarce or where export of tissues is a hampering factor. While there are studies that have assessed the usefulness of this sequencing device in potential field settings (Mulley and Hargreaves 2015), we could find relatively few examples where the MinION had been used in remote field sites such as those experienced by many ornithologists. One of the most common uses of the MinION in remote locations has been in a biomedical context, including real-time genomic surveillance of the Ebola virus in Africa (Hoenen et al. 2016; Quick et al. 2016) and the Zika virus in Brazil (Faria et al. 2016). Incorporating this technology in the surveillance of avian zoonotic diseases, for example AIV, could facilitate rapid identification of the virus causing an outbreak in birds, and hence help to make a quick decision

on the action required to minimise the spread of infection. While these studies proved the utility of the MinION at various isolated locations, they all depended on a minimal set of laboratory equipment such as micro-centrifuges and thermocyclers for sample processing, and a stable Internet connection for data analysis (Faria et al. 2016; Hoenen et al. 2016; Quick et al. 2016), all of which could be e.g. fitted into a camping car making it a mobile mini-laboratory. These obstacles need to be overcome before the MinION can be used routinely in the field. We believe that the continuous development of efficient library preparation kits as well as tools enabling data analyses offline will make the MinION a realistic choice for field transcriptomics in the future.

Acknowledgements Open access funding provided by Max Planck Society. We are thankful to Dr Anna Cederlund for her illustrations. We apologise if we have missed references on other non-model avian whole-transcriptome studies.

References

Abu-Bonsrah KD, Zhang D, Newgreen DF (2016) CRISPR/Cas9 targets chicken embryonic somatic cells in vitro and in vivo and generates phenotypic abnormalities. Sci Rep 6:34524. https://doi.org/10.1038/srep34524

Afrakhte M, Schultheiss TM (2004) Construction and analysis of a subtracted library and microarray of cDNAs expressed specifically in chicken heart progenitor cells. Dev Dyn 230:290–298. https://doi.org/10.1002/dvdy.20059

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. Mol Biol 215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Ålund née Podevin M (2017) Sex, sperm and speciation: on sexual selection and fertility in hybridizing flycatchers. Dissertation, Acta Universitatis Upsaliensis

Alvarez M, Schrey AW, Richards CL (2015) Ten years of transcriptomics in wild populations: what have we learned about their ecology and evolution? Mol Ecol 24:710–725. https://doi.org/10.1111/mec.13055

Anders S, Huber W (2010) Differential expression analysis for sequence count data. Genome Biol 11:R106. https://doi.org/10.1186/gb-2010-11-10-r106

Andersson L et al (2015) Coordinated international action to accelerate genome-to-phenome with FAANG, the functional annotation of animal genomes project. Genome Biol 16:57. https://doi.org/10.1186/s13059-015-0622-4

Annabelle T, Karine R, Marie-Dominique B, Stéphane D, Karine G (2017) Kinetics of expression of genes involved in glucose metabolism after the last meal in overfed Mule Ducks. Mol Cell Biochem 430:127. https://doi.org/10.1007/s11010-017-2960-x

Arai E et al (2017) Physiological conditions and genetic controls of phaeomelanin pigmentation in nestling Barn Swallows. Behav Ecol 28:706–716. https://doi.org/10.1093/beheco/arx012

Ast G (2004) How did alternative splicing evolve? Nat Rev Genet 5:773–782. https://doi.org/10.1038/nrg1451

Ayub M, Hardwick SW, Luisi BF, Bayley H (2013) Nanopore-based identification of individual nucleotides for direct RNA sequencing. Nano Lett 13:6144–6150. https://doi.org/10.1021/nl403469r

Backström N, Zhang Q, Edwards SV (2013) Evidence from a House Finch (*Haemorhous mexicanus*) spleen transcriptome for adaptive evolution and biased gene conversion in passerine birds. Mol Biol Evol 30:1046–1050. https://doi.org/10.1093/molbev/mst033

Balakrishnan CN, Edwards SV, Clayton DF (2010) The Zebra Finch genome and avian genomics in the wild. Emu 110:233–241. https://doi.org/10.1071/MU09087

Balakrishnan CN, Lin Y-C, London SE, Clayton DF (2012) RNA-seq transcriptome analysis of male and female Zebra Finch cell lines. Genomics 100:363–369. https://doi.org/10.1016/j.ygeno.2012.08.002

Balakrishnan CN, Chapus C, Brewer MS, Clayton DF (2013) Brain transcriptome of the Violet-eared Waxbill *Uraeginthus granatina* and recent evolution in the songbird genome. Open Biol 3:130063. https://doi.org/10.1098/rsob.130063

Balakrishnan CN, Mukai M, Gonser RA, Wingfield JC, London SE, Tuttle EM, Clayton DF (2014) Brain transcriptome sequencing and assembly of three songbird model systems for the study of social behavior. PeerJ 2:e396. https://doi.org/10.7717/peerj.396

Bao S, Jiang R, Kwan W, Wang B, Ma X, Song Y-Q (2011) Evaluation of next-generation sequencing software in mapping and assembly. J Hum Genet 56:406–414. https://doi.org/10.1038/jhg.2011.43

Bartel DP (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 116:281–297. https://doi.org/10.1016/S0092-8674(04)00045-5

Baverstock P, Adams M, Polkinghorne R, Gelder M (1982) A sex-linked enzyme in birds—Z-chromosome conservation but no dosage compensation. Nature 296:763–766. https://doi.org/10.1038/296763a0

Bélteky J, Agnvall B, Jensen P (2017) Gene expression of behaviorally relevant genes in the cerebral hemisphere changes after selection for tameness in Red Junglefowl. PLoS ONE 12:e0177004. https://doi.org/10.1371/journal.pone.0177004

Bliss TW, Dohms JE, Emara MG, Keeler CL (2005) Gene expression profiling of avian macrophage activation. Vet Immunol Immunopathol 105:289–299. https://doi.org/10.1016/j.vetimm.2005.02.013

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bowlin MS et al (2010) Grand challenges in migration biology. Integr Comp Biol 50:261–279. https://doi.org/10.1093/icb/icq013

Brookes AJ (1999) The essence of SNPs. Gene 234:177–186. https://doi.org/10.1016/S0378-1119(99)00219-X

Bullard JH, Purdom E, Hansen KD, Dudoit S (2010) Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. BMC Bioinf. 11:94. https://doi.org/10.1186/1471-2105-11-94

Burnside J et al (2005) Development of a cDNA array for chicken gene expression analysis. BMC Genomics 6:1. https://doi.org/10.1186/1471-2164-6-13

Busby MA, Stewart C, Miller CA, Grzeda KR, Marth GT (2013) Scotty: a web tool for designing RNA-Seq experiments to measure differential gene expression. Bioinformatics 29:656–657. https://doi.org/10.1093/bioinformatics/btt015

Bustin SA et al (2009) The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. Clin Chem 55:611–622. https://doi.org/10.1373/clinchem.2008

Cantarel BL, Korf I, Robb SMC, Parra G, Ross E, Moore B, Holt C, Sanchez Alvarado A, Yandell M (2007) MAKER: an easy-to-use

annotation pipeline designed for emerging model organism genomes. Genome Res 18(1):188–196
Cao N, Li W, Li B, Tian Y, Xu D (2017) Transcriptome profiling reveals the immune response of goose T cells under selenium stimuli. Anim Sci J. https://doi.org/10.1111/asj.12861
Capecchi MR (2005) Gene targeting in mice: functional analysis of the mammalian genome for the twenty-first century. Nat Rev Genet 6:507–512. https://doi.org/10.1038/nrg1619
Chapman JR, Waldenström J (2015) With reference to reference genes: a systematic review of endogenous controls in gene expression studies. PLoS ONE 10:e0141853. https://doi.org/10.1371/journal.pone.0141853
Charlesworth B (1996) The evolution of chromosomal sex determination and dosage compensation. Curr Biol 6:149–162. https://doi.org/10.1016/S0960-9822(02)00448-7
Chen C-K et al (2016) Regulatory differences in natal down development between altricial Zebra Finch and precocial chicken. Mol Biol Evol 33:msw085. https://doi.org/10.1093/molbev/msw085
Chen C-K et al (2017a) Identification and evolutionary analysis of long non-coding RNAs in Zebra Finch. BMC Genomics 18:117. https://doi.org/10.1186/s12864-017-3506-z
Chen S et al (2017b) Goose Mx and Oasl play vital roles in the antiviral effects of type I, II, and III interferon against newly emerging avian Flavivirus. Front Immunol 8:1006. https://doi.org/10.3389/fimmu.2017.01006
Cheviron Z, Swanson D (2017) Comparative transcriptomics of seasonal phenotypic flexibility in two North American songbirds. Integr Comp Biol 57:1040–1054. https://doi.org/10.1093/icb/icx118
Cheviron ZA, Whitehead A, Brumfield RT (2008) Transcriptomic variation and plasticity in Rufous-collared Sparrows (*Zonotrichia capensis*) along an altitudinal gradient. Mol Ecol 17:4556–4569. https://doi.org/10.1111/j.1365-294X.2008.03942.x
Cheviron ZA, Carling MD, Brumfield RT (2011) Effects of postmortem interval and preservation method on RNA isolated from field-preserved avian tissues. Condor 113:483–489. https://doi.org/10.1525/cond.2011.100201
Chiu C, Miller S (2016) Next-generation sequencing. In: Persing DH, Tenover FC, Hayden RT, Leven G, Miller MB, Nolte FS (eds) Molecular microbiology: diagnostic principles and practice, 3rd edn. ASM, Washington, pp 68–79. https://doi.org/10.1128/9781555819071.ch6
Chu Y, Corey DR (2012) RNA sequencing: platform selection, experimental design, and data interpretation. Nucleic Acid Ther 22:271–274. https://doi.org/10.1089/nat.2012.0367
Chu J-H, Lin R-C, Yeh C-F, Hsu Y-C, Li S-H (2012) Characterization of the transcriptome of an ecologically important avian species, the Vinous-throated Parrotbill *Paradoxornis webbianus bulomachus* (Paradoxornithidae; Aves). BMC Genomics 13:149. https://doi.org/10.1186/1471-2164-13-149
Chung O et al (2015) The first whole genome and transcriptome of the Cinereous Vulture reveals adaptation in the gastric and immune defense systems and possible convergent evolution between the Old and New World vultures. Genome Biol 16:1–11. https://doi.org/10.1186/s13059-015-0780-4
Clancy S (2008) RNA functions. Nat Educ 1:102
Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21:3674–3676
Conesa A et al (2016) A survey of best practices for RNA-seq data analysis. Genome Biol 17:13. https://doi.org/10.1186/s13059-016-0881-8
Cooke TF et al (2017) Genetic mapping and biochemical basis of yellow feather pigmentation in Budgerigars. Cell 171:427–439. e421. https://doi.org/10.1016/j.cell.2017.08.016
Cooper CA et al (2017) Generation of gene edited birds in one generation using sperm transfection assisted gene editing (STAGE). Transgenic Res 26:331–347. https://doi.org/10.1007/s11248-016-0003-0
Core Team R (2014) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, p 2013
Costa V, Aprile M, Esposito R, Ciccodicola A (2013) RNA-Seq and human complex diseases: recent accomplishments and future perspectives. Eur J Hum Genet 21:134–142. https://doi.org/10.1038/ejhg.2012.129
Crosetto N, Bienko M, van Oudenaarden A (2015) Spatially resolved transcriptomics and beyond. Nat Rev Genet 16:57–66. https://doi.org/10.1038/nrg3832
Dalloul RA et al (2010) Multi-platform next-generation sequencing of the Domestic Turkey (*Meleagris gallopavo*): genome assembly and analysis. PLoS Biol 8:e1000475. https://doi.org/10.1371/journal.pbio.1000475
Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nat Rev Genet 12:499–510. https://doi.org/10.1038/nrg3012
Davidson JH, Balakrishnan CN (2016) Gene regulatory evolution during speciation in a songbird. G3 Genes Genom Genet 6:1357–1364. https://doi.org/10.1534/g3.116.027946
Davidson NM, Oshlack A (2014) Corset: enabling differential gene expression analysis for de novo assembled transcriptomes. Genome Biol 15:410. https://doi.org/10.1186/s13059-014-0410-6
De Wit P et al (2012) The simple fool's guide to population genomics via RNA-Seq: an introduction to high-throughput sequencing data analysis. Mol Ecol Resour 12:1058–1067. https://doi.org/10.1111/1755-0998.12003
De Wit P, Pespeni MH, Palumbi SR (2015) SNP genotyping and population genomics from expressed sequences–current advances and future possibilities. Mol Ecol 24:2310–2323. https://doi.org/10.1111/mec.13165
Derveaux S, Vandesompele J, Hellemans J (2010) How to do successful gene expression analysis using real-time PCR. Methods 50:227–230. https://doi.org/10.1016/j.ymeth.2009.11.001
Désert C et al (2016) Transcriptomes of whole blood and PBMC in chickens. Comp Biochem Phys D 20:1–9. https://doi.org/10.1016/j.cbd.2016.06.008
Dick MF (2017) The long haul: migratory flight preparation and performance in songbirds. Dissertation, the University of Western Ontario
Dingle H, Drake VA (2007) What is migration? Bioscience 57:113–121. https://doi.org/10.1641/B570206
Dotti I, Bonin S (2011) DNase treatment of RNA. In: Stanta G (ed) Guidelines for molecular analysis in archive tissues. Springer, Berlin, pp 87–90
Doudna JA, Charpentier E (2014) The new frontier of genome engineering with CRISPR-Cas9. Science 346:1258096. https://doi.org/10.1126/science.1258096
Eberwine J, Sul J-Y, Bartfai T, Kim J (2014) The promise of single-cell sequencing. Nat Methods 11:25–27. https://doi.org/10.1038/nmeth.2769
Ekblom R, Galindo J (2011) Applications of next generation sequencing in molecular ecology of non-model organisms. Heredity 107:1–15. https://doi.org/10.1038/hdy.2010.152
Ekblom R, Wang B (2017) Development of transcriptome genetic markers for the Great Snipe (*Gallinago media*). Conserv Genet Resour. https://doi.org/10.1007/s12686-017-0746-9
Ekblom R, Balakrishnan CN, Burke T, Slate J (2010a) Digital gene expression analysis of the Zebra Finch genome. BMC Genomics 11:219. https://doi.org/10.1186/1471-2164-11-219

Ekblom R, French L, Slate J, Burke T (2010b) Evolutionary analysis and expression profiling of Zebra Finch immune genes. Genome Biol Evol 2:781–790. https://doi.org/10.1093/gbe/evq061

Ekblom R, Wennekes P, Horsburgh GJ, Burke T (2014) Characterization of the House Sparrow (*Passer domesticus*) transcriptome: a resource for molecular ecology and immunogenetics. Mol Ecol Resour 14:636–646. https://doi.org/10.1111/1755-0998.12213

Ellegren H (2000) Evolution of the avian sex chromosomes and their role in sex determination. Trends Ecol Evol 15:88–192. https://doi.org/10.1016/s0169-5347(00)01821-8

Ellegren H (2004) Microsatellites: simple sequences with complex evolution. Nat Rev Genet 5:435–445. https://doi.org/10.1038/nrg1348

Ellegren H (2008) Sequencing goes 454 and takes large-scale genomics into the wild. Mol Ecol 17:1629–1631. https://doi.org/10.1111/j.1365-294X.2008.03699.x

Ellegren H (2011) Sex-chromosome evolution: recent progress and the influence of male and female heterogamety. Nat Rev Genet 12:157–166. https://doi.org/10.1038/nrg2948

Ellegren H, Sheldon BC (2008) Genetic basis of fitness differences in natural populations. Nature 452:169–175. https://doi.org/10.1038/nature06737

Ellegren H et al (2012) The genomic landscape of species divergence in *Ficedula* flycatchers. Nature 491:756–760. https://doi.org/10.1038/nature11584

Espina V et al (2006) Laser-capture microdissection. Nat Protoc 1:586–603. https://doi.org/10.1038/nprot.2006.85

Faria NR, Sabino EC, Nunes MR, Alcantara LCJ, Loman NJ, Pybus OG (2016) Mobile real-time surveillance of Zika virus in Brazil. Genome Med 8:97. https://doi.org/10.1186/s13073-016-0356-2

Farlie PG et al (2017) Co-option of the cardiac transcription factor Nkx2. 5 during development of the Emu wing. Nat Commun. https://doi.org/10.1038/s41467-017-00112-7

Finseth FR, Harrison RG (2014) A comparison of next generation sequencing technologies for transcriptome assembly and utility for RNA-Seq in a non-model bird. PLoS ONE 9:e108550. https://doi.org/10.1371/journal.pone.0108550

Finseth FR, Harrison RG (2017) Genes integral to the reproductive function of male reproductive tissues drive heterogeneity in evolutionary rates in Japanese Quail. G3 Genes Genom Genet. https://doi.org/10.1534/g3.117.300095

Fleming-Canepa X et al (2011) Expression of duck CCL19 and CCL21 and CCR7 receptor in lymphoid and influenza-infected tissues. Mol Immunol 48:1950–1957. https://doi.org/10.1016/j.molimm.2011.05.025

Franchini P, Irisarri I, Fudickar A, Schmidt A, Meyer A, Wikelski M, Partecke J (2017) Animal tracking meets migration genomics: transcriptomic analysis of a partially migratory bird species. Mol Ecol 26:3204–3216. https://doi.org/10.1111/mec.14108

Freedman ML et al (2011) Principles for the post-GWAS functional characterization of cancer risk loci. Nat Genet 43:513–518. https://doi.org/10.1038/ng.840

Fudickar AM, Peterson MP, Greives TJ, Atwell JW, Bridge ES, Ketterson ED (2016) Differential gene expression in seasonal sympatry: mechanisms involved in diverging life histories. Biol Lett 12:20160069. https://doi.org/10.1098/rsbl.2016.0069

Garalde DR et al (2016) Highly parallel direct RNA sequencing on an array of nanopores. bioRxiv. https://doi.org/10.1101/068809

Gene Ontology Consortium (2004) The gene ontology (GO) database and informatics resource. Nucleic Acids Res 32:D258–D261. https://doi.org/10.1093/nar/gkh036

Gill F, Donsker D Eds (2017) IOC World Bird List (version 7.3). https://doi.org/10.14344/ioc.ml.7.3

Gish W (1993) Identification of protein-coding regions by database similarity search. Nat Genet 3:266–272. https://doi.org/10.1038/ng0393-266

Goecks J, Nekrutenko A, Taylor J, Team TG (2010) Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. Genome Biol 11:R86. https://doi.org/10.1186/gb-2010-11-8-r86

Gonzalez-Garay ML (2016) Introduction to isoform sequencing using pacific biosciences technology (Iso-Seq). In: Wu J (ed) Transcriptomics and gene regulation. Springer, Dordrecht, pp 141–160. https://doi.org/10.1007/978-94-017-7450-5_6

Grabherr MG et al (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol 29:644–652. https://doi.org/10.1038/nbt.1883

Gunaratne PH et al (2011) Song exposure regulates known and novel microRNAs in the Zebra Finch auditory forebrain. BMC Genomics 12:277. https://doi.org/10.1186/1471-2164-12-277

Haas BJ, Zody MC (2010) Advancing RNA-seq analysis. Nat Biotechnol 28:421–423. https://doi.org/10.1038/nbt0510-421

Hagen IJ, Billing AM, Rønning B, Pedersen SA, Pärn H, Slate J, Jensen H (2013) The easy road to genome-wide medium density SNP screening in a non-model species: development and application of a 10 K SNP-chip for the House Sparrow (*Passer domesticus*). Mol Ecol Resour 13:429–439. https://doi.org/10.1111/1755-0998.12088

Han B, Li Y, Han H, Zhao Y, Pan Q, Ren L (2017) Three IgH isotypes, IgM, IgA and IgY are expressed in Gentoo Penguin and Zebra Finch. PLoS ONE 12:e0173334. https://doi.org/10.1371/journal.pone.0173334

Harrington CA, Rosenow C, Retief J (2000) Monitoring gene expression using DNA microarrays. Curr Opin Microbiol 3:285–291. https://doi.org/10.1016/S1369-5274(00)00091-6

Harrison PW, Wright AE, Zimmer F, Dean R, Montgomery SH, Pointer MA, Mank JE (2015) Sexual selection drives evolution and rapid turnover of male gene expression. Proc Natl Acad Sci USA 112:4393–4398. https://doi.org/10.1073/pnas.1501339112

Hartl DL, Clark A (2007) Principles of population genetics. Sinauer, Sunderland

Hashimshony T, Wagner F, Sher N, Yanai I (2012) CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. Cell Rep 2:666–673. https://doi.org/10.1016/j.celrep.2012.08.003

Hedenström A (2008) Adaptations to migration in birds: behavioural strategies, morphology and scaling effects. Philos Trans R Soc B 363:287–299. https://doi.org/10.1098/rstb.2007.2140

Hillier LW et al (2004) Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature 432:695–716. https://doi.org/10.1038/nature03154

Hoenen T et al (2016) Nanopore sequencing as a rapidly deployable Ebola outbreak tool. Emerg Infect Dis 22:331. https://doi.org/10.3201/eid2202.151796

Höglund J et al (2017) Blood transcriptomes and de novo identification of candidate loci for mating success in lekking Great Snipe (*Gallinago media*). Mol Ecol 26:3458–3471. https://doi.org/10.1111/mec.14118

Huang Y et al (2013) The duck genome and transcriptome provide insight into an avian influenza virus reservoir species. Nat Genet 45:776–783. https://doi.org/10.1038/ng.2657

Ilinskaya O, Mahmud RS (2014) Ribonucleases as antiviral agents. Mol Biol 48:615–623. https://doi.org/10.1134/S0026893314040050

Jaitin DA et al (2014) Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. Science 343:776–779. https://doi.org/10.1126/science.1247651

Johnston RA, Paxton KL, Moore FR, Wayne RK, Smith TB (2016) Seasonal gene expression in a migratory songbird. Mol Ecol 25:5680–5691. https://doi.org/10.1111/mec.13879

Josefsen MH, Löfström C, Hansen T, Reynisson E, Hoorfar J (2012) Instrumentation and fluorescent chemistries used in quantitative polymerase chain reaction. In: Filion M (ed) Quantitative

real-time PCR in applied microbiology. Caister, Norfolk, pp 27–52
Junker JP et al (2014) Genome-wide RNA tomography in the Zebrafish embryo. Cell 159:662–675. https://doi.org/10.1016/j.cell.2014.09.038
Kaiser SA, Taylor SA, Chen N, Sillett TS, Bondra ER, Webster MS (2017) A comparative assessment of SNP and microsatellite markers for assigning parentage in a socially monogamous bird. Mol Ecol Resour 17:183–193. https://doi.org/10.1111/1755-0998.12589
Kanehisa M, Goto S (2000) KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res 28:27–30. https://doi.org/10.1093/nar/28.1.27
Kanehisa M, Goto S, Furumichi M, Tanabe M, Hirakawa M (2010) KEGG for representation and analysis of molecular networks involving diseases and drugs. Nucleic Acids Res 38:D355–D360. https://doi.org/10.1093/nar/gkp896
King M-C, Wilson AC (1975) Evolution at two levels in humans and chimpanzees. Science 188:107–116
Koglin S, Trense D, Wink M, Sauer-Gürth H, Tietze DT (2017) Characterization of a de novo assembled transcriptome of the Common Blackbird (*Turdus merula*). PeerJ 5:e4045. https://doi.org/10.7717/peerj.4045
Kolodziejczyk AA, Kim JK, Svensson V, Marioni JC, Teichmann SA (2015) The technology and biology of single-cell RNA sequencing. Mol Cell 58:610–620. https://doi.org/10.1016/j.molcel.2015.04.005
Korlach J et al (2017) De novo PacBio long-read and phased avian genome assemblies correct and add to reference genes generated with intermediate and short reads. GigaScience 6:1–16. https://doi.org/10.1093/gigascience/gix085
Kozarewa I, Ning Z, Quail MA, Sanders MJ, Berriman M, Turner DJ (2009) Amplification-free Illumina sequencing-library preparation facilitates improved mapping and assembly of (G + C)-biased genomes. Nat Methods 6:291–295. https://doi.org/10.1038/nmeth.1311
Kratochwil CF, Meyer A (2015) Closing the genotype–phenotype gap: emerging technologies for evolutionary genetics in ecological model vertebrate systems. BioEssays 37:213–226. https://doi.org/10.1002/bies.201400142
Kratz A, Carninci P (2014) The devil in the details of RNA-seq. Nat Biotechnol 32:882–884. https://doi.org/10.1038/nbt.3015
Kraus RH, Wink M (2015) Avian genomics: fledging into the wild! J Ornithol 156:851–865. https://doi.org/10.1007/s10336-015-1253-y
Kraus RH, Van Hooft P, Waldenström J, Latorre-Margalef N, Ydenberg R (2009) Avian influenza surveillance: on the usability of FTA cards to solve biosafety and transport issues. Wildfowl 2:215–223
Kraus RH, van Hooft P, Waldenström J, Latorre-Margalef N, Ydenberg RC, Prins HH (2011) Avian influenza surveillance with FTA cards: field methods, biosafety, and transportation issues solved. Jove J Vis Exp 54:e2832. https://doi.org/10.3791/2832
Künstner A et al (2010) Comparative genomics based on massive parallel transcriptome sequencing reveals patterns of substitution and selection across 10 bird species. Mol Ecol 19:266–276. https://doi.org/10.1111/j.1365-294X.2009.04487.x
Künstner A, Nabholz B, Ellegren H (2011) Evolutionary constraint in flanking regions of avian genes. Mol Biol Evol 28:2481–2489. https://doi.org/10.1093/molbev/msr066
Kuo RI, Tseng E, Eory L, Paton IR, Archibald AL, Burt DW (2017) Normalized long read RNA sequencing in chicken reveals transcriptome complexity similar to human. BMC Genomics 18:323. https://doi.org/10.1186/s12864-017-3691-9
Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359. https://doi.org/10.1038/nmeth.1923
Law CW, Alhamdoosh M, Su S, Smyth GK, Ritchie ME (2016) RNA-seq analysis is easy as 1-2-3 with limma, Glimma and edgeR. F1000Res 5:1408. https://doi.org/10.12688/f1000research.9005.2
Le Duc D et al (2015) kiwi genome provides insights into evolution of a nocturnal lifestyle. Genome Biol 16:147. https://doi.org/10.1186/s13059-015-0711-4
Leng N et al (2013) EBSeq: an empirical Bayes hierarchical model for inference in RNA-seq experiments. Bioinformatics 29:1035–1043. https://doi.org/10.1093/bioinformatics/btt087
Levin JZ et al (2010) Comprehensive comparative analysis of strand-specific RNA sequencing methods. Nat Methods 7:709–715. https://doi.org/10.1038/nmeth.1491
Li B, Dewey CN (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12:323. https://doi.org/10.1186/1471-2105-12-323
Li X, Chiang H-I, Zhu J, Dowd SE, Zhou H (2008) Characterization of a newly developed chicken 44 K Agilent microarray. BMC Genomics 9:60. https://doi.org/10.1186/1471-2164-9-60
Li Z et al (2015) MicroRNAs in the immune organs of chickens and ducks indicate divergence of immunity against H5N1 avian influenza. FEBS Lett 589:419–425. https://doi.org/10.1016/j.febslet.2014.12.019
Liu Y, Zhou J, White KP (2013) RNA-seq differential expression studies: more sequence or more replication? Bioinformatics 30:301–304
Lockhart DJ, Winzeler EA (2000) Genomics, gene expression and DNA arrays. Nature 405:827–836. https://doi.org/10.1038/35015701
Looi Q, Amin H, Aini I, Zuki M, Omar A (2017) De novo transcriptome analysis shows differential expression of genes in salivary glands of edible bird's nest producing swiftlets. BMC Genomics 18:504. https://doi.org/10.1186/s12864-017-3861-9
Lopes RJ et al (2016) Genetic basis for red coloration in birds. Curr Biol 26:1427–1434. https://doi.org/10.1016/j.cub.2016.03.076
Lorkowski S, Cullen PM (2006) Analysing gene expression, a handbook of methods: possibilities and pitfalls. Wiley, Weinheim
Love MI, Anders S, Huber W (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15:550. https://doi.org/10.1186/s13059-014-0550-8
Lu L et al (2015) The goose genome sequence leads to insights into the evolution of waterfowl and susceptibility to fatty liver. Genome Biol 16:89. https://doi.org/10.1186/s13059-015-0652-y
Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of population genomics: from genotyping to genome typing. Nat Rev Genet 4:981–994. https://doi.org/10.1038/nrg1226
Lundberg M et al (2013) Characterisation of a transcriptome to find sequence differences between two differentially migrating subspecies of the Willow Warbler *Phylloscopus trochilus*. BMC Genomics 14:330. https://doi.org/10.1186/1471-2164-14-330
MacManes MD, Austin SH, Lang AS, Booth A, Farrar V, Calisi RM (2017) Widespread patterns of sexually dimorphic gene expression in an avian hypothalamic–pituitary–gonadal (HPG) axis. Sci Rep 7:45125. https://doi.org/10.1038/srep45125
Macosko EZ et al (2015) Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. Cell 161:1202–1214. https://doi.org/10.1016/j.cell.2015.05.002
Mamanova L et al (2010) FRT-seq: amplification-free, strand-specific, transcriptome sequencing. Nat Methods 7:130. https://doi.org/10.1038/nmeth.1417
Mank JE (2009) Sex chromosomes and the evolution of sexual dimorphism: lessons from the genome. Am Nat 173:141–150. https://doi.org/10.1086/595754
Mank JE, Vicoso B, Berlin S, Charlesworth B (2010) Effective population size and the faster-X effect: empirical results and their

interpretation. Evolution 64:663–674. https://doi.org/10.1111/j.1558-5646.2009.00853.x
Martin LB, Coon CA, Liebl AL, Schrey AW (2014) Surveillance for microbes and range expansion in House Sparrows. Proc R Soc B Biol Sci 281:20132690. https://doi.org/10.1098/rspb.2013.2690
Mazzoni G, Kadarmideen HN (2016) Computational methods for quality check, preprocessing and normalization of RNA-seq data for systems biology and analysis. In: Kadarmideen HN (ed) Systems biology in animal production and health, vol 2. Springer, Cham, pp 61–77. https://doi.org/10.1007/978-3-319-43332-5_3
McDonald PG, Griffith SC (2011) To pluck or not to pluck: the hidden ethical and scientific costs of relying on feathers as a primary source of DNA. J Avian Biol 42:197–203. https://doi.org/10.1111/j.1600-048X.2011.05365.x
McKenna A et al (2010) The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20:1297–1303. https://doi.org/10.1101/gr.107524.110
Meisel RP, Connallon T (2013) The faster-X effect: integrating theory and data. Trends Genet 29:537–544. https://doi.org/10.1016/j.tig.2013.05.009
Meitern R, Andreson R, Hõrak P (2014) Profile of whole blood gene expression following immune stimulation in a wild passerine. BMC Genomics 15:533. https://doi.org/10.1186/1471-2164-15-533
Metzker ML (2010) Sequencing technologies—the next generation. Nat Rev Genet 11:31–46. https://doi.org/10.1038/nrg2626
Modrek B, Lee C (2002) A genomic view of alternative splicing. Nat Genet 30:13–19. https://doi.org/10.1038/ng0102-13
Moon DA, Veniamin SM, Parks-Dely JA, Magor KE (2005) The MHC of the duck (*Anas platyrhynchos*) contains five differentially expressed class I genes. J Immunol 175:6702–6712. https://doi.org/10.4049/jimmunol.175.10.6702
Morin PA, Luikart G, Wayne RK (2004) SNPs in ecology, evolution and conservation. Trends Ecol Evol 19:208–216. https://doi.org/10.1016/j.tree.2004.01.009
Morozova O, Hirst M, Marra MA (2009) Applications of new sequencing technologies for transcriptome analysis. Annu Rev Genom Hum G 10:135–151. https://doi.org/10.1146/annurev-genom-082908-145957
Morris KV, Mattick JS (2014) The rise of regulatory RNA. Nat Rev Genet 15:423–437. https://doi.org/10.1038/nrg3722
Mueller JC, Kuhl H, Timmermann B, Kempenaers B (2015) Characterization of the genome and transcriptome of the Blue Tit *Cyanistes caeruleus*: polymorphisms, sex-biased expression and selection signals. Mol Ecol Resour 16:549–561. https://doi.org/10.1111/1755-0998.12450
Müller GB (2007) Evo–devo: extending the evolutionary synthesis. Nat Rev Genet 8:943–949. https://doi.org/10.1038/nrg2219
Mulley JF, Hargreaves AD (2015) Snake venom gland cDNA sequencing using the Oxford nanopore MinION portable DNA sequencer. bioRxiv. https://doi.org/10.1101/025148
Nabholz B, Jarvis ED, Ellegren H (2010) Obtaining mtDNA genomes from next-generation transcriptome sequencing: a case study on the basal Passerida (Aves: Passeriformes) phylogeny. Mol Phylogenet Evol 57:466–470. https://doi.org/10.1016/j.ympev.2010.06.009
Nabholz B, Künstner A, Wang R, Jarvis ED, Ellegren H (2011) Dynamic evolution of base composition: causes and consequences in avian phylogenomics. Mol Biol Evol 28:2197–2210. https://doi.org/10.1093/molbev/msr047
Naurin S et al (2008) TECHNICAL ADVANCES: a microarray for large-scale genomic and transcriptional analyses of the Zebra Finch (*Taeniopygia guttata*) and other passerines. Mol Ecol Resour 8:275–281. https://doi.org/10.1111/j.1471-8286.2007.01979.x
Naurin S, Hasselquist D, Bensch S, Hansson B (2012) Sex-biased gene expression on the avian Z chromosome: highly expressed genes show higher male-biased expression. PLoS ONE 7:e46854. https://doi.org/10.1371/journal.pone.0046854
Necsulea A, Kaessmann H (2014) Evolutionary dynamics of coding and non-coding transcriptomes. Nat Rev Genet 15:734–748. https://doi.org/10.1038/nrg3802
Neiman PE et al (2001) Analysis of gene expression during myc oncogene-induced lymphomagenesis in the bursa of Fabricius. Proc Natl Acad Sci USA 98:6378–6383. https://doi.org/10.1073/pnas.111144898
Newhouse DJ, Hofmeister EK, Balakrishnan CN (2017) Transcriptional response to West Nile virus infection in the Zebra Finch (*Taeniopygia guttata*). R Soc Open Sci 4:170296. https://doi.org/10.1098/rsos.170296
Nielsen R, Paul JS, Albrechtsen A, Song YS (2011) Genotype and SNP calling from next-generation sequencing data. Nat Rev Genet 12:443–451. https://doi.org/10.1038/nrg2986
Nimpf S et al (2017) Subcellular analysis of pigeon hair cells implicates vesicular trafficking in cuticulosome formation and maintenance. eLife. https://doi.org/10.7554/eLife.29959
Noble WS (2009) How does multiple testing correction work? Nat Biotechnol 27:1135–1137. https://doi.org/10.1038/nbt1209-1135
Nolan T, Hands RE, Bustin SA (2006) Quantification of mRNA using real-time RT-PCR. Nat Protoc 1:1559–1582. https://doi.org/10.1038/nprot.2006.236
Ockendon NF et al (2016) Optimization of next-generation sequencing transcriptome annotation for species lacking sequenced genomes. Mol Ecol Resour 16:446–458. https://doi.org/10.1111/1755-0998.12465
Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M (1999) KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res 27:29–34. https://doi.org/10.1093/nar/27.1.29
Ohno S (1959) Sex chromosomes and sex-linked genes. Teratology 4:111
Oishi I, Yoshii K, Miyahara D, Kagami H, Tagami T (2016) Targeted mutagenesis in chicken using CRISPR/Cas9 system. Sci Rep 6:23980. https://doi.org/10.1038/srep23980
Oleksiak MF, Churchill GA, Crawford DL (2002) Variation in gene expression within and among natural populations. Nat Genet 32:261–266. https://doi.org/10.1038/ng983
Oshlack A, Robinson MD, Young MD (2010) From RNA-seq reads to differential expression results. Genome Biol 11:220. https://doi.org/10.1186/gb-2010-11-12-220
Oyler-McCance SJ, Oh KP, Langin KM, Aldridge CL (2016) A field ornithologist's guide to genomics: practical considerations for ecology and conservation. Auk 133:626–648. https://doi.org/10.1642/AUK-16-49.1
Ozsolak F, Milos PM (2011) RNA sequencing: advances, challenges and opportunities. Nat Rev Genet 12:87–98. https://doi.org/10.1038/nrg2934
Ozsolak F et al (2009) Direct RNA sequencing. Nature 461:814–818. https://doi.org/10.1038/nature08390
Pan S et al (2017) Population transcriptomes reveal synergistic responses of DNA polymorphism and RNA expression to extreme environments on the Qinghai-Tibetan Plateau in a predatory bird. Mol Ecol 26:2993–3010. https://doi.org/10.1111/mec.14090
Pantalacci S, Sémon M (2015) Transcriptomics of developing embryos and organs: a raising tool for evo–devo. J Exp Zool B 324:363–371. https://doi.org/10.1002/jez.b.22595
Park TS, Lee HJ, Kim KH, Kim J-S, Han JY (2014) Targeted gene knockout in chickens mediated by TALENs. Proc Natl Acad Sci USA 111:12716–12721. https://doi.org/10.1073/pnas.1410555111

Parsch J, Ellegren H (2013) The evolutionary causes and consequences of sex-biased gene expression. Nat Rev Genet 14:83–87. https://doi.org/10.1038/nrg3376

Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL (2016) Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nat Protoc 11:1650–1667. https://doi.org/10.1038/nprot.2016.095

Peterson MP et al (2012) De novo transcriptome sequencing in a songbird, the Dark-eyed Junco (*Junco hyemalis*): genomic tools for an ecological model system. BMC Genomics 13:305. https://doi.org/10.1186/1471-2164-13-305

Pfaffl MW (2012) Quantification strategies in real-time polymerase chain reaction. In: Filion M (ed) Quantitative real-time PCR in applied microbiology. Caister, Norfolk, pp 53–62

Picelli S, Björklund ÅK, Faridani OR, Sagasser S, Winberg G, Sandberg R (2013) Smart-seq2 for sensitive full-length transcriptome profiling in single cells. Nat Methods 10:1096–1098. https://doi.org/10.1038/nmeth.2639

Quick J et al (2016) Real-time, portable genome sequencing for Ebola surveillance. Nature 530:228–232. https://doi.org/10.1038/nature16996

Ramstad KM, Miller HC, Kolle G (2016) Sixteen kiwi (*Apteryx* spp.) transcriptomes provide a wealth of genetic markers and insight into sex chromosome evolution in birds. BMC Genomics 17:410. https://doi.org/10.1186/s12864-016-2714-2

Rao JR, Fleming CC, Moore JE (2006) Molecular diagnostics: current technology and applications. Horizon Bioscience, Norfolk

Rhoads A, Au KF (2015) PacBio sequencing and its applications. Genomics Proteomics Bioinformatics 13:278–289. https://doi.org/10.1016/j.gpb.2015.08.002

Richardson MF, Sherwin WB, Rollins LA (2017) De novo assembly of the liver transcriptome of the European Starling, *Sturnus vulgaris*. J Genomics 5:54–57. https://doi.org/10.7150/jgen.19504

Robinson MD, Oshlack A (2010) A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol 11:R25. https://doi.org/10.1186/gb-2010-11-3-r25

Sandford EE et al (2012) Strong concordance between transcriptomic patterns of spleen and peripheral blood leukocytes in response to avian pathogenic *Escherichia coli* infection. Avian Dis 56:732–736. https://doi.org/10.1637/10261-060512-Reg.1

Santure AW, Gratten J, Mossman JA, Sheldon BC, Slate J (2011) Characterisation of the transcriptome of a wild Great Tit *Parus major* population by next generation sequencing. BMC Genomics 12:753–770. https://doi.org/10.1186/1471-2164-12-283

Schatz MC, Langmead B, Salzberg SL (2010) Cloud computing and the DNA data race. Nat Biotechnol 28:691–693. https://doi.org/10.1038/nbt0710-691

Schena M, Shalon D, Davis RW, Brown PO (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 270:467

Schlötterer C (2004) The evolution of molecular markers—Just a matter of fashion? Nat Rev Genet 5:63–69. https://doi.org/10.1038/nrg1249

Schulze A, Downward J (2001) Navigating gene expression using microarrays: a technology review. Nat Cell Biol 3:E190–E195. https://doi.org/10.1038/35087138

Schurch NJ et al (2016) How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use? RNA 22:839–851. https://doi.org/10.1261/rna.053959.115

Schusser B et al (2013) Immunoglobulin knockout chickens via efficient homologous recombination in primordial germ cells. Proc Natl Acad Sci USA 110:20170–20175. https://doi.org/10.1073/pnas.1317106110

Selkoe KA, Toonen RJ (2006) Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. Ecol Lett 9:615–629. https://doi.org/10.1111/j.1461-0248.2006.00889.x

Sevane N, Cañon J, Eusebi PG, Gil I, Dunner S (2017) Red-legged Partridge (*Alectoris rufa*) de-novo transcriptome assembly and identification of gene-related markers. Genomics Data 11:132–134. https://doi.org/10.1016/j.gdata.2017.02.003

Seyednasrollah F, Laiho A, Elo LL (2015) Comparison of software packages for detecting differential expression in RNA-seq studies. Brief Bioinform 16:59–70. https://doi.org/10.1093/bib/bbt086

Shukla RN (2015) Analysis of chromosome. Agrotech, Jaipur

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31:3210–3212. https://doi.org/10.1093/bioinformatics/btv351

Smith J et al (2006) Development of a chicken 5 K microarray targeted towards immune function. BMC Genomics 7:49. https://doi.org/10.1186/1471-2164-7-49

Smith J et al (2015) A comparative analysis of host responses to avian influenza infection in ducks and chickens highlights a role for the interferon-induced transmembrane proteins in viral resistance. BMC Genomics 16:574. https://doi.org/10.1186/s12864-015-1778-8

Soneson C, Delorenzi M (2013) A comparison of methods for differential expression analysis of RNA-seq data. BMC Bioinformatics 14:91. https://doi.org/10.1186/1471-2105-14-91

Spurgeon SL, Jones RC, Ramakrishnan R (2008) High throughput gene expression measurement with real time PCR in a microfluidic dynamic array. PLoS ONE 3:e1662. https://doi.org/10.1371/journal.pone.0001662

Srivastava A (2011) Evolution & detection of non-coding RNA, and transcriptome analyses of two non-model systems. Dissertation, University of Georgia

Srivastava A, Winker K, Shaw TI, Jones KL, Glenn TC (2012) Transcriptome analysis of a North American songbird, *Melospiza melodia*. DNA Res 19:325–333. https://doi.org/10.1093/dnares/dss015

Ståhl PL et al (2016) Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. Science 353:78–82. https://doi.org/10.1126/science.aaf2403

Subramanian S, Huynen L, Millar C, Lambert D (2010) Next generation sequencing and analysis of a conserved transcriptome of New Zealand's kiwi. BMC Evol Biol 10:387. https://doi.org/10.1186/1471-2148-10-387

Swanson DL, Sabirzhanov B, VandeZande A, Clark TG (2009) Seasonal variation of myostatin gene expression in pectoralis muscle of House Sparrows (*Passer domesticus*) is consistent with a role in regulating thermogenic capacity and cold tolerance. Physiol Biochem Zool 82:121–128. https://doi.org/10.1086/591099

Takekawa JY, Hill NJ, Schultz AK, Iverson SA, Cardona CJ, Boyce WM, Dudley JP (2011) Rapid diagnosis of avian influenza virus in wild birds: use of a portable rRT-PCR and freeze-dried reagents in the field. J Vis Exp. https://doi.org/10.3791/2829

Tang F et al (2009) mRNA-Seq whole-transcriptome analysis of a single cell. Nat Methods 6:377–382. https://doi.org/10.1038/nmeth.1315

Tariq M, Chen R, Yuan H, Liu Y, Wu Y, Wang J, Xia C (2015) De novo transcriptomic analysis of peripheral blood lymphocytes from the Chinese Goose: gene discovery and immune system pathway description. PLoS ONE 10:e0121015. https://doi.org/10.1371/journal.pone.0121015

Tautz D, Ellegren H, Weigel D (2010) Next generation molecular ecology. Mol Ecol 19:1–3. https://doi.org/10.1111/j.1365-294X.2009.04489.x

Teng M et al (2016) A benchmark for RNA-seq quantification pipelines. Genome Biol 17:74. https://doi.org/10.1186/s13059-016-0940-1

The IUCN Red List of Threatened Species. Version 2017-3 (2017) www.iucnredlist.org. Accessed 13 Dec 2017

Todd EV, Black MA, Gemmell NJ (2016) The power and promise of RNA-seq in ecology and evolution. Mol Ecol 25:1224–1241. https://doi.org/10.1111/mec.13526

Toews DP et al (2015) Genomic approaches to understanding population divergence and speciation in birds. Auk 133:13–30. https://doi.org/10.1642/AUK-15-51.1

Tomfohr J, Lu J, Kepler TB (2005) Pathway level analysis of gene expression using singular value decomposition. BMC Bioinformatics 6:225. https://doi.org/10.1186/1471-2105-6-225

Trapnell C, Salzberg SL (2009) How to map billions of short reads onto genomes. Nat Biotechnol 27:455–457. https://doi.org/10.1038/nbt0509-455

Uebbing S, Künstner A, Mäkinen H, Ellegren H (2013) Transcriptome sequencing reveals the character of incomplete dosage compensation across multiple tissues in flycatchers. Genome Biol Evol 5:1555–1566. https://doi.org/10.1093/gbe/evt114

Uebbing S et al (2016) Divergence in gene expression within and between two closely related flycatcher species. Mol Ecol 25:2015–2028. https://doi.org/10.1111/mec.13596

Van Den Bussche RA, Judkins ME, Montague MJ, Warren WC (2017) A resource of genome-wide single nucleotide polymorphisms (Snps) for the conservation and management of Golden Eagles. J Raptor Res 51:368–377. https://doi.org/10.3356/JRR-16-47.1

Van Goor A, Ashwell CM, Persia ME, Rothschild MF, Schmidt CJ, Lamont SJ (2017) Unique genetic responses revealed in RNA-seq of the spleen of chickens stimulated with lipopolysaccharide and short-term heat. PLoS ONE 12:e0171414. https://doi.org/10.1371/journal.pone.0171414

Van Hemert S, Ebbelaar BH, Smits MA, Rebel JM (2003) Generation of EST and microarray resources for functional genomic studies on chicken intestinal health. Anim Biotechnol 14:133–143. https://doi.org/10.1081/ABIO-120026483

VanGuilder HD, Vrana KE, Freeman WM (2008) Twenty-five years of quantitative PCR for gene expression analysis. Biotechniques 44:619. https://doi.org/10.2144/000112776

Videvall E, Cornwallis CK, Palinauskas V, Valkiūnas G, Hellgren O (2015) The avian transcriptome response to malaria infection. Mol Biol Evol 32:1255–1267. https://doi.org/10.1093/molbev/msv016

Videvall E, Cornwallis CK, Ahrén D, Palinauskas V, Valkiūnas G, Hellgren O (2017) The transcriptome of the avian malaria parasite *Plasmodium ashfordi* displays host-specific gene expression. Mol Ecol 26:2939–2958. https://doi.org/10.1111/mec.14085

Vijay N, Poelstra JW, Künstner A, Wolf JB (2013) Challenges and strategies in transcriptome assembly and differential gene expression quantification. A comprehensive in silico assessment of RNA-seq experiments. Mol Ecol 22:620–634. https://doi.org/10.1111/mec.12014

Vijayakumar P, Raut AA, Kumar P, Sharma D, Mishra A (2014) De novo assembly and analysis of crow lungs transcriptome. Genome 57:499–506. https://doi.org/10.1139/gen-2014-0122

Vijayakumar P, Mishra A, Ranaware PB, Kolte AP, Kulkarni DD, Burt DW, Raut AA (2015) Analysis of the crow lung transcriptome in response to infection with highly pathogenic H5N1 avian influenza virus. Gene 559:77–85. https://doi.org/10.1016/j.gene.2015.01.016

Wada K et al (2006) A molecular neuroethological approach for identifying and characterizing a cascade of behaviorally regulated genes. Proc Natl Acad Sci USA 103:15212–15217. https://doi.org/10.1073/pnas.0607098103

Wade J et al (2004) A cDNA microarray from the telencephalon of juvenile male and female Zebra Finches. J Neurosci Meth 138:199–206. https://doi.org/10.1016/j.jneumeth.2004.04.007

Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 10:57–63. https://doi.org/10.1038/nrg2484

Wang Y, Ghaffari N, Johnson CD, Braga-Neto UM, Wang H, Chen R, Zhou H (2011) Evaluation of the coverage and depth of transcriptome by RNA-Seq in chickens. BMC Bioinformatics 12:S5. https://doi.org/10.1186/1471-2105-12-s10-s5

Wang Y et al (2012) Integrated analysis of microRNA expression and mRNA transcriptome in lungs of avian influenza virus infected broilers. BMC Genomics 13:278. https://doi.org/10.1186/1471-2164-13-278

Wang Z, Zhang J, Yang W, An N, Zhang P, Zhang G, Zhou Q (2014) Temporal genomic evolution of bird sex chromosomes. BMC Evol Biol 14:250. https://doi.org/10.1186/s12862-014-0250-8

Wang M, Uebbing S, Ellegren H (2015a) Allele-specific gene expression inferred by a Bayesian negative binomial approach indicates abundant *cis*-regulatory variation in natural flycatcher populations. In: Uebbing S (ed) On the evolution of the avian transcriptome. Dissertation. University of Uppsala, Uppsala

Wang Q, Wu Y, Cai Y, Zhuang Y, Xu L, Baocheng W, Zhang Y (2015b) Spleen transcriptome profile of ducklings in response to infection with Muscovy Duck reovirus. Avian Dis 59:282–290. https://doi.org/10.1637/10992-112514-Reg

Wang M, Uebbing S, Ellegren H (2017a) Bayesian Inference of allele-specific gene expression indicates abundant *cis*-regulatory variation in natural flycatcher populations. Genome Biol Evol 9:1266–1279. https://doi.org/10.1093/gbe/evx080

Wang Q et al (2017b) Transcriptomic analysis reveals the molecular mechanism of apoptosis induced by Muscovy Duck reovirus. Genes Genom. https://doi.org/10.1007/s13258-017-0567-y

Warren WC et al (2010) The genome of a songbird. Nature 464:757–762. https://doi.org/10.1038/nature08819

Watson H, Videvall E, Andersson MN, Isaksson C (2017) Transcriptome analysis of a wild bird reveals physiological responses to the urban environment. Sci Rep 7:44180. https://doi.org/10.1038/srep44180

Westermann AJ, Gorski SA, Vogel J (2012) Dual RNA-seq of pathogen and host. Nat Rev Microbiol 10:618–630. https://doi.org/10.1038/nrmicro2852

Whitehead A, Crawford DL (2006) Variation within and among species in gene expression: raw material for evolution. Mol Ecol 15:1197–1211. https://doi.org/10.1111/j.1365-294X.2006.02868.x

Wilhelm BT, Landry J-R (2009) RNA-Seq—quantitative measurement of expression through massively parallel RNA-sequencing. Methods 48:249–257. https://doi.org/10.1016/j.ymeth.2009.03.016

Wolf JB (2013) Principles of transcriptome analysis and gene expression quantification: an RNA-seq tutorial. Mol Ecol Resour 13:559–572. https://doi.org/10.1111/1755-0998.12109

Wolf JB, Bryk J (2011) General lack of global dosage compensation in ZZ/ZW systems? Broadening the perspective with RNA-seq. BMC Genomics 12:91. https://doi.org/10.1186/1471-2164-12-91

Wolf JB, Bayer T, Haubold B, Schilhabel M, Rosenstiel P, Tautz D (2010) Nucleotide divergence vs. gene expression differentiation: comparative transcriptome sequencing in natural isolates from the Carrion Crow and its hybrid zone with the Hooded Crow. Mol Ecol 19:162–175. https://doi.org/10.1111/j.1365-294X.2009.04471.x

Workman RE, Myrka AM, Tseng E, Wong GW, Welch KC, Timp W (2017) Single molecule, full-length transcript sequencing provides insight into the extreme metabolism of Ruby-throated Hummingbird *Archilochus colubris*. bioRxiv. https://doi.org/10.1101/117218

Wright AE, Harrison PW, Montgomery SH, Pointer MA, Mank JE (2014) Independent stratum formation on the avian sex chromosomes reveals inter-chromosomal gene conversion and

predominance of purifying selection on the W chromosome. Evolution 68:3281–3295. https://doi.org/10.1111/evo.12493
Wright AE, Harrison PW, Zimmer F, Montgomery SH, Pointer MA, Mank JE (2015) Variation in promiscuity and sexual selection drives avian rate of Faster-Z evolution. Mol Ecol 24:1218–1235. https://doi.org/10.1111/mec.13113
Xu X et al (2016) Transcriptomic analysis of different stages of pigeon ovaries by RNA-sequencing. Mol Reprod Dev 83:640–648. https://doi.org/10.1002/mrd.22670
Yandell M, Ence D (2012) A beginner's guide to eukaryotic genome annotation. Nat Rev Genet 13:329–342. https://doi.org/10.1038/nrg3174
Young MD, Wakefield MJ, Smyth GK, Oshlack A (2010) Gene ontology analysis for RNA-seq: accounting for selection bias. Genome Biol 11:R14. https://doi.org/10.1186/gb-2010-11-2-r14
Zhan X et al (2013) Peregrine and saker falcon genome sequences provide insights into evolution of a predatory lifestyle. Nat Genet 45:563–566. https://doi.org/10.1038/ng.2588
Zhang G (2015) Genomics: bird sequencing project takes off. Nature 522:34. https://doi.org/10.1038/522034d
Zhang G, Jarvis ED, Gilbert MTP (2014a) A flock of genomes. Science 346:1308–1309. https://doi.org/10.1126/science.346.6215.1308
Zhang Q, Hill GE, Edwards SV, Backström N (2014b) A House Finch (*Haemorhous mexicanus*) spleen transcriptome reveals intra-and interspecific patterns of gene expression, alternative splicing and genetic diversity in passerines. BMC Genomics 15:305. https://doi.org/10.1186/1471-2164-15-305
Zhao S, Fung-Leung W-P, Bittner A, Ngo K, Liu X (2014) Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. PLoS ONE 9:e78644. https://doi.org/10.1371/journal.pone.0078644
Zhen Y et al (2017) Genomic divergence across ecological gradients in the Central African Rainforest Songbird (*Andropadus virens*). Mol Ecol 26:4966–4977. https://doi.org/10.1111/mec.14270
Zhou Q et al (2014) Complex evolutionary trajectories of sex chromosomes across bird taxa. Science 346:1246338. https://doi.org/10.1126/science.1246338