CHAPTER TEN

High-Throughput Sequencing Will Metamorphose the Analysis of Thyroid Hormone Receptor Function During Amphibian Development

Alexis G. Grimaldi*, Nicolas Buisine*, Patrice Bilesimo*,1, Laurent M. Sachs*,2

*Laboratoire d'Evolution des Regulations Endocrinienne, Muséum National d'Histoire Naturelle, Département Régulation, Développement et Diversité Moléculaire, UMR 7221 CNRS, Paris, France
[1]Present address: Translational Research Center, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA
[2]Corresponding author: e-mail address: sachs@mnhn.fr

Contents

Abstract

Amphibian metamorphosis is marked by dramatic thyroid hormone (T_3)-induced changes including *de novo* morphogenesis, tissue remodeling, and organ resorption through programmed cell death. These changes involve cascades of gene regulation initiated by thyroid hormone (TH). TH functions by regulating gene expression through TH receptors (TR). TR are DNA-binding transcription factors that belong to the steroid hormone receptor superfamily. In the absence of ligand, TR can repress gene expression by recruiting a corepressor complex, whereas liganded TR recruits a coactivator

Current Topics in Developmental Biology, Volume 103
ISSN 0070-2153
http://dx.doi.org/10.1016/B978-0-12-385979-2.00010-1

complex for gene activation. Earlier studies have led us to propose a dual function model for TR during development. In premetamorphic tadpoles, unliganded TR represses transcription involving corepressors. During metamorphosis, endogenous T_3 allows TR to activate gene expression. To fully understand the diversity of T_3 effects during metamorphosis, whole genome analysis of transcriptome and mechanism of TR action should be carried out. To this end, the new sequencing technologies have dramatically changed how fundamental questions in biology are being addressed and is now making the transition from technology development to being a standard for genomic and functional genomic analysis. This review focuses on the applications of high-throughput technologies to the field of amphibian metamorphosis.

1. INTRODUCTION

Amphibian development is marked by an indirect development. Embryonic development and juvenile life style are separated by a larval period (tadpole) that ends with metamorphosis. Metamorphosis is a switch that results in the reprogramming of the morphological and biochemical characteristics of nearly all tadpole organs (Shi, 1999). Transformations involve apoptosis of larval cells and concurrent proliferation and differentiation of adult cell types. Although complex, metamorphosis is directly initiated by thyroid hormone (TH) with 3,5,3′-triiodothyronine (T_3) as the most biologically active agent. Interestingly, this hormone-controlled developmental phase shows strong similarities with the perinatal period in mammals, bird hatching, and of course other kind of metamorphosis such as in fish and insects (Laudet, 2011). Amphibian metamorphosis thus provides a powerful model to analyze the essential roles for TH, notably in maturation of the nervous system, the metabolism, the respiratory system, the intestine, and body shape modification (Shi, 1999; Tata, 1993).

TH acts through thyroid hormone receptor (TR) that belongs to the family of nuclear receptors (Gronemeyer & Laudet, 1995). TR is a transcription factor that regulates gene expression by binding to DNA at specific sites known as T_3 response elements (T_3REs), mostly composed of two direct repeats of the consensus sequence "AGGTCA" separated by four bases (DR4). TH is a versatile player. It can not only upregulate, but also downregulate gene expression in target tissues or cells (Flamant, Gauthier, & Samarut, 2007). Moreover, in the absence of TH, TR can also repress or increase the transcription. As in mammals, there are two types of TR (α and β) in amphibian (Yaoita, Shi, & Brown, 1990). Their amino acid sequences are well conserved in evolution and amphibian TR behaves

similarly to the mammalian or avian TR. Unfortunately, there are only a very limited number of studies on gene repression by liganded TR (Santos, Fairall, & Schwabe, 2011). Thus, most discussions have been focused on the mechanism for T_3-response genes induced by T_3 (Grimaldi, Buisine, Miller, Shi, & Sachs, 2012). For these genes, in absence of hormone, TR binds T_3RE and recruits corepressor complexes carrying chromatin modification activity that leads to a closed chromatin conformation, inaccessible for the transcriptional machinery. T_3-binding induces a conformational change of TR that will lead first, to the release of the corepressor complexes and second, to the recruitment of coactivator complexes carrying a different set of chromatin modification enzymes. Their actions lead to chromatin opening favorable to transcription activation.

The expression profiles (Yaoita & Brown, 1990), together with the ability of TR to both repress and activate T_3-inducible genes in the absence and presence of TH (Sachs & Shi, 2000), respectively, suggest dual functions for TRs during development (Sachs et al., 2000). TRα and TRβ mRNA are present at very low levels during embryogenesis where their function has started to be addressed (Havis et al., 2006). Next, high levels of TRα mRNA are present after larval hatching and increase to reach a maximum during metamorphosis. TRβ mRNA levels are barely detectable before metamorphosis, when they will increase in parallel with endogenous T_3 level, again reaching a peak at the climax. After metamorphosis, the levels of both TR isoforms decrease to stay low in the juvenile and the adult. The same profiles were observed analyzing protein levels (Eliceiri & Brown, 1994). The tissue distribution of TR has shown to be correlated with organ transformation (Eliceiri & Brown, 1994; Fairclough & Tata, 1997; Kawahara, Baker, & Tata, 1991). Thus, in premetamorphic tadpoles, TRs, mostly TRα, act to repress T_3-response genes. As many of these genes are likely to participate in metamorphosis, their repression by unliganded TR will help prevent premature metamorphosis and ensure a proper period of tadpole growth. TH synthesis and secretion into the plasma transforms TRs from repressors to activators, which will induce the expression of TH response genes leading to metamorphosis.

Analysis of TH action in amphibian metamorphosis will provide a powerful tool to understand the diversity of biological activities known for TH and the mechanisms that lead to this diversity. Studies in the past two decades using subtractive hybridization or DNA array have identified many T_3-response genes during amphibian metamorphosis (Brown et al., 1996; Buchholz, Heimeier, Das, Washington, & Shi, 2007; Buckbinder &

Brown, 1992; Cai, Das, & Brown, 2007; Das et al., 2006; Das, Heimeier, Buchholz, & Shi, 2009; Shi & Brown, 1993; Wang & Brown, 1991, 1993). Among them, only a small fraction has been characterized as direct response gene (Das et al., 2009; Furlow & Brown, 1999; Ranjan, Wong, & Shi, 1994). The identification of the TH-induced transcriptional regulatory programs needs now to be addressed in a physiological context at the level of the whole genome to understand how TH regulates metamorphosis. Such ambitious studies depend on the exploitation of a sequenced amphibian genome. While *Xenopus leavis (X. laevis)* is the most used model, *Xenopus tropicalis (X. tropicalis)* is by far a more appropriate model because its genome sequence is in a large part already accessible (Hellsten et al., 2010). Now next-generation sequencing (NGS)-based transcriptome sequencing, transcription factor binding, and epigenetic analyses (chromatin modification) have dramatically changed how fundamental questions in biology are addressed (Hawkins, Hon, & Ren, 2010; Metzker, 2010; Sboner, Mu, Greenbaum, Auerbach, & Gerstein, 2011). NGS is well suited to provide all the primary data required for functional genomic analyses with an unprecedented resolution enabling the detection of even subtle differences. This review emphasizes the particular contribution of NGS-based technologies to functional genomics research with a special focus on gene regulation by TH.

2. *XENOPUS TROPICALIS* GENOMIC SEQUENCE AND ANNOTATION: CURRENT STATE

2.1. A fragmented genome assembly

The sequencing of *X. tropicalis* genome has been an ongoing project for nearly 10 years. It was initiated by the JGI in late 2003, and the first public version (3.0) was released in October 2004. The draft assembly version 4.1 released in 2005 had been in use for over 5 years before its publication (Hellsten et al., 2010). A new assembly (version 7.1) quickly followed and was made available to download on the XenBase Web site in 2011(Bowes et al., 2008). These assemblies have been an invaluable resource for the scientific community. In fact, as opposed to *X. laevis*, which is pseudotetraploid and has a genome of about 3.1 Gbp, *X. tropicalis* has a relatively small genome (1.7 Gbp) and much less genetic redundancy. This is expected since it is a diploid species, making it a very appropriate model to probe, for example, the TH-induced gene regulatory networks. This

general purpose genomic resource has been instrumental in key recent findings of *X. tropicalis* biology (Bowes et al., 2008; Hellsten et al., 2010).

Unlike mouse and human genomes, *X. tropicalis* genome project did not benefit from the inputs and curations of a large and extensive scientific community. Indeed, this is the case for almost all eukaryotes genomes published so far, except for a few model organisms. As a result, the draft genome assembly v4.1 of *X. tropicalis* is composed of 19,501 scaffolds (see Fig. 10.1A; Buisine & Sachs, 2009) representing unordered fragments of the 10 chromosomes (Tymowska & Fischberg, 1973). In the version 4.1, the largest scaffold is 7.8 Mb long and the N50 = 1.6 Mb. The version 7.1 is quite less fragmented, with only 7730 scaffolds, an N50 of 130 Mb, with the longest scaffold being 216 Mb long. Interestingly, a small number of the largest scaffolds capture a majority of the genes: 80% of the Ensembl genes in the 834 top scaffolds (Fig. 10.1A). In addition, many assembly gaps (stretches of "N"s corresponding to unsequenced regions) are found in scaffolds and add an extra level of fragmentation (Fig. 10.1B and C). Given that many multiexon genes are several tens of kilobases long, such gaps will frequently be found in gene loci, resulting in alteration of apparent distances (Fig. 10.1B), as well as missing exons or gene fragments. Thus, although most of the genome sequence is reflected by a small subset of scaffolds (80% of the assembly in the top 2000 scaffolds), fragmentation remains high (Fig. 10.1C). BAC, cosmid and fosmid libraries have been end-sequenced to resolve the physical continuity over large size range (Hellsten et al., 2010). Nonetheless, end-sequenced scaffolds are still present in the assembly, artificially increasing the size of the assembled genome by ~85 Mbp. They cluster as dense clumps of points on Fig. 10.1A (highlighted by the three leftmost ellipses). In addition, some regions (end-sequences) may be represented multiple times in the genome sequence. While the assembled genome is 1.5 Gbp long, one has to take into account that overall, 20% of the released assembly is actually not sequenced. However, unassembled regions of the genome assembly result from low complexity regions, or repeated sequences, which makes both base detection—in the case of long stretches of similar nucleotides—and assembly—because of the ambiguity introduced—more difficult. In addition, assembly gaps could also be due to the presence of allelic polymorphism, which could induce hard to resolve ambiguities during the assembly process. Intraspecies polymorphism can be ruled out in the case of *X. tropicalis* since only one inbred individual frog was used for genome sequencing (Hellsten et al., 2010).

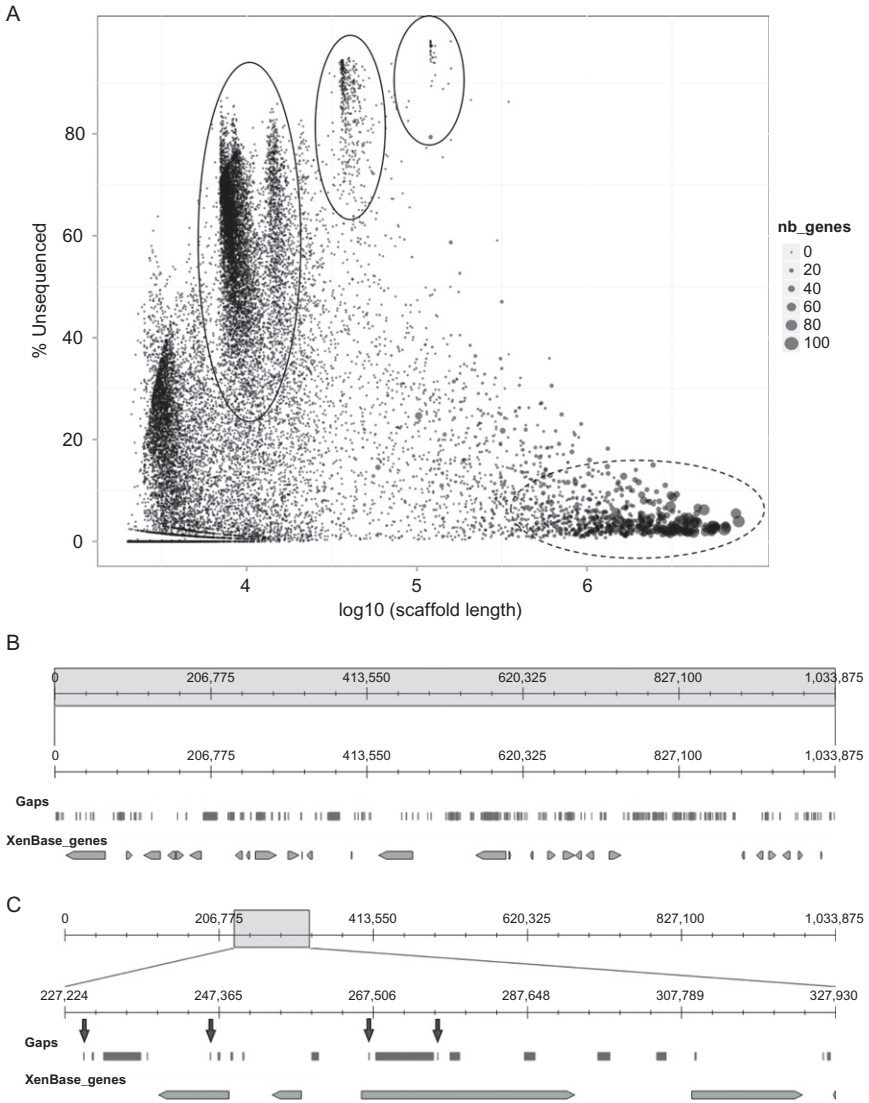

Figure 10.1 The *X. tropicalis* genome assembly 4.1 is highly fragmented. (A) Scatterplot representing the size of each scaffold (in $\log_{10}$ scale) compared to the proportion of assembly gaps (% Unsequenced). The gene number per scaffold (ranging from 0 to 100) is indicated by gray circles of various sizes. The three ellipses in the upper half of the plot correspond to a fraction of the cosmid, fosmid, and BAC-end libraries used to help the assembly process. The gene rich and almost fully sequenced scaffolds capturing ~80% of the assembly are indicated at the bottom-right of the plot (dashed-circle). (B) Illustrative example of the assembly gaps found in a scaffold. The first track shows the assembly gaps (stretches of 'N's). The second track represents XenBase gene models. The orientation of each gene is indicated by the direction of the arrow. The scaffold shown is scaffold_448. (C) Close-up of a 100 kb region of the scaffold_448. Assembly gaps that are exactly 50 bp long ("gaps of 50," see text) are indicated by arrows.

A specific subset of assembly gaps find their origin in a technical choice made during the assembly process. Indeed, nearly half of the ~170,000 assembly gaps are exactly 50 bp long. They were in fact arbitrarily introduced when the length of an unsequenced region could not be determined. They are heterogeneously distributed across the genome and can be found both inside of genes as well as in intergenic regions (Fig. 10.1C). The use of the gPET technology (Fullwood, Wei, Liu, & Ruan, 2009) suggest that the size of these "gaps of 50" is in fact underestimated by at least an order of magnitude, although the individual size of each gap is subject to large variations (unpublished data). Such inconsistencies can have a dramatic impact in the exploitation of the genome sequence as a general purpose resource by a biologist unaware of these pitfalls, since regions apparently close together can actually be located several kilobases apart from each other. This is not much of a problem when studying a small region of the genome which can be visually inspected (in the range of a few hundred base pairs up to a few kilobases), but this can become a real issue when large genomic regions are being analyzed. Furthermore, annotated features as short as 50 bp do not show up well in genome browsers when displaying genomic regions larger than 10 kbs, which deprives the biologist of visual clues for regions containing "gaps of 50." Of note, this point has been partially corrected in the version 7.1, although 42% 47,505 assembly gaps are exactly 100 bp and 10 kb.

2.2. An incomplete annotation

Despite the fragmentation level of the genome assembly, the majority of genes are included in the longest scaffolds (Fig. 10.1A, bottom-right circled cluster), which supports the idea that the current assembly contains most of the gene rich (euchromatic) regions. Many different gene annotation datasets are available for *X. tropicalis:* Ensembl (Flicek et al., 2012), XenBase (Bowes et al., 2008), and RefSeq (Pruitt, Tatusova, Klimke, & Maglott, 2009). They are more or less comprehensive, redundant, and complementing each other. For example, while Ensembl has more gene models than XenBase (18,025 vs. 15,929 out of more than 22,000 expected), it also has more models with no associated name ("unknown," 6293 vs. 1074). In fact, as shown in Fig. 10.2A, the three databases provide gene models that are quite comparable qualitatively, but clearly lack consensus in the definition of their boundaries. Indeed, the operational value of a genome sequence is (in part) dictated by the quality of its structural and functional annotations.

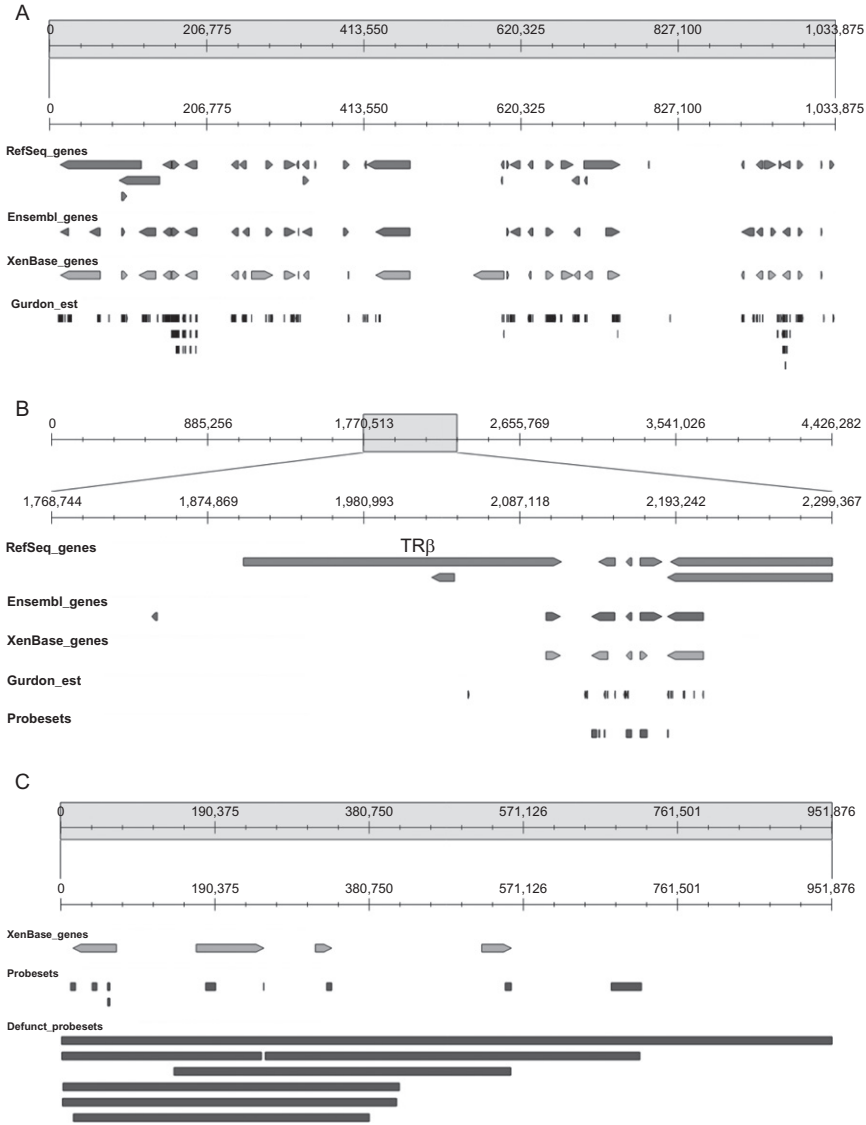

Figure 10.2 Limits of the current gene annotation and microarray design. (A) Comparison gene annotations from different sources. The first three tracks represent gene models described in RefSeq, Ensembl, and XenBase. The orientation of each gene model is indicated by arrows. The fourth track corresponds to the EST clusters computed at the Gurdon institute (available on the XenBase website). The scaffold shown is scaffold_448. (B) Affymetrix probesets located around the TRβ locus (scaffold_26). The four top tracks correspond to the same datasets as in A. The last track represents the genomic region spanned by the Affymetrix *X. tropicalis* transcriptome microarrays probesets, curated from defunct probesets. Note that there are no probeset associated with TRβ. Thus,

Precise definition of gene boundaries is of utmost importance when trying to link coordinated variation of gene expression with the specific recruitment of transcription factors or modification of the chromatin architecture. In particular, most of the current gene models capture only the coding sequence (CDS), without taking into account the untranslated regions (UTRs). The upstream boundary of the 5′UTR defines the actual transcription start site (TSS) and proximal promoter region, with which many regulatory elements are associated and which are required for transcriptional regulation. The same is true for 3′UTRs, which are underrepresented and often contain numerous microRNA-binding sites, which may be important posttranscriptional regulators.

Traditionally, the definition of gene models relies on the use of a collection of softwares (*ab initio* gene prediction, comparison to databases, etc.) and functional data (ESTs, full-length cDNA, etc.). Between 2000 and 2007, the number of EST for *X. tropicalis* has increased tremendously, reaching more than 1.2 million ESTs (grouped by unigene library entry, ∼5 million altogether), making it one of the model with the highest number of ESTs. The ESTs were clustered and aligned to the genome sequence in order to define potential exons (Nagaraj, Gasser, & Ranganathan, 2007). Although this strategy can be quite effective when large collections of EST are used, the process remains error prone (depending on the quality of the collections) and often lacks the resolution required to describe alternative transcripts (Nagaraj et al., 2007). As can be seen in Fig. 10.2B, the gene TRβ is associated to a single EST cluster far upstream of the annotated TSS in Ensembl or XenBase, while only the RefSeq annotation reflects boundaries that actually encompass the EST cluster.

Gene models can be inferred from the genomic sequence using computational methods. Popular gene prediction softwares include genewise (Birney, Clamp, & Durbin, 2004), geneMark, FGENESH (Salamov & Solovyev, 2000), and EUGENE (Schiex, Moisan, & Rouzé, 2001), to name a few. This process suffers several pitfalls and always benefits from multiple lines of evidence. In addition, assembled genome must be of relatively good

TRβ transcription cannot be measured with Affymetrix arrays. (C) Illustrative example of defunct Affymetrix probesets. The defunct probesets are composed of probes poorly specific of a given gene/transcript. The first track corresponds to the XenBase gene models as in (A). The second track corresponds to genomic region spanned by the curated probesets that can be used for analysis, while the third track (defunct_probesets) corresponds to the genomic region spanned by defunct probesets. The region shown is located on scaffold_464.

quality, as unsequenced portions as well as assembly errors can lead to missing or truncated gene models. Also, the information content along the genome can be extremely variable, in such a way that pseudogenes, duplications, short tandem repeats, and the overall complexity level of a region can mislead the software. Many misannotated gaps inside the gene body, as shown in Fig. 10.1C, can further reduce the efficiency of gene prediction. Furthermore, predicted gene models are often limited to the CDS, and miss UTRs which, as previously stated, often contain important regulatory elements. In *X. tropicalis* genome, some key genes involved in metamorphosis or TH signaling are poorly annotated, which can hamper downstream analysis of functional data. For example, based on the current annotation, it would be impossible to find the T_3RE located at the TSS of TRβ by looking for it near the 5′ end of the gene (a very common practice, although many binding sites may be missed as they are located away from the TSS, see below). In fact, the 5′ end of the TRβ gene was determined by 5′ RACE PCR and is correctly annotated only in the manually curated RefSeq database. It is located 210 kb upstream of the XenBase and Ensembl models (Fig. 10.2B).

Gene models can be determined by full cDNA sequencing (Klein et al., 2002; Voigt, Chen, Gilchrist, Amaya, & Papalopulu, 2005). Although this method is quite expensive and slow, it is also very accurate. Individual cDNA clones are assembled from multiple single-pass sequenced reads, ensuring the whole coverage of the cloned molecules. When combined, these approaches can yield accurate models, although extensive curation is still required. While there is an expected ∼22,000 protein coding genes, only half are actually linked to accession numbers pointing to cDNA in the RefSeq database, and only one-third are linked to full-length cDNA (Gilchrist, 2012).

The lack of clear definition of gene boundaries renders the analysis of coordinated variation of gene expression quite tricky. In fact, genome-wide measure of differential gene expression has traditionally been carried out using microarrays, whose design relies heavily on a good annotation (Altmann et al., 2001). In this technology, fluorescent labeled cDNA is hybridized to probes fixed on a solid surface. The readout of the signal for a given spot then depends on the relative amount of cDNA present in the sample. On the chip, transcript-specific probes define probesets, with the redundancy of the individual probes allowing the assay to be more specific and sensitive. The design of microarrays thus requires a comprehensive annotation, and the quality and exhaustiveness of the chip will greatly depend on it. In model organisms (mouse, drosophila, etc.), microarrays have been

constantly refined, and inconsistent probes were detected in the numerous datasets available in GEO and Array Express and then excluded from analyses. In *X. tropicalis*, the number of microarray datasets available in the databases has not reached the critical value needed to do such curation, and large portion of the probes present on available microarrays are not usable, as they are located outside of exons/genes, or in misannotated regions. As a result, probesets (i.e., a collection of probes aimed at measuring the expression level of a single transcript or gene) often span over large genomic regions, irrespective of the underlying gene definition provided by the annotation (Fig. 10.2C). In addition, after filtering of the defunct probesets, the remaining probesets do not interrogate the whole transcriptome. For example, some genes critically involved in metamorphosis, like TRβ (Fig. 10.2B), are missing in the probesets.

3. TH AND TR REGULATING NETWORK USING NGS-BASED TRANSCRIPTOME ANALYSIS

3.1. Toward a better annotation

Traditional resources used to annotate genomes are either low-throughput (in the case of EST and full-length cDNAs) or error prone (in the case of *de novo*, *in silico* predictions) and they are combined to take the best of both and increase specificity and sensitivity of gene prediction methods (see above). The recent rise of NGS technologies and their applications to RNA sequencing (RNA-Seq) opens new perspectives in term of genome annotation, given their very high throughput, scalability, and robustness (Wang, Gerstein, & Snyder, 2009). RNA-Seq is reminiscent of EST sequencing, except that the sequencing step is carried out by an NGS system rather than a conventional capillary sequencing system. To this end, cDNAs are synthesized by reverse transcription from oligo-dT purified messenger RNAs, and sheared in smaller fragments suitable for sequencing. With regular RNA-Seq, only one end of each fragment of sheared cDNAs is sequenced over a short region, typically $\leq$50 bp. Sequence reads are then mapped onto the genome with one of the many short reads aligners. This simple protocol is best suited for quantitative measure of gene expression, as the normalized number of reads per gene is proportional to its expression level, but can also provide a quick description of transcription units (Garber, Grabherr, Guttman, & Trapnell, 2011). This is well illustrated in the case of the TRβ gene (Fig. 10.3A), in which the RNA-Seq signal can be detected

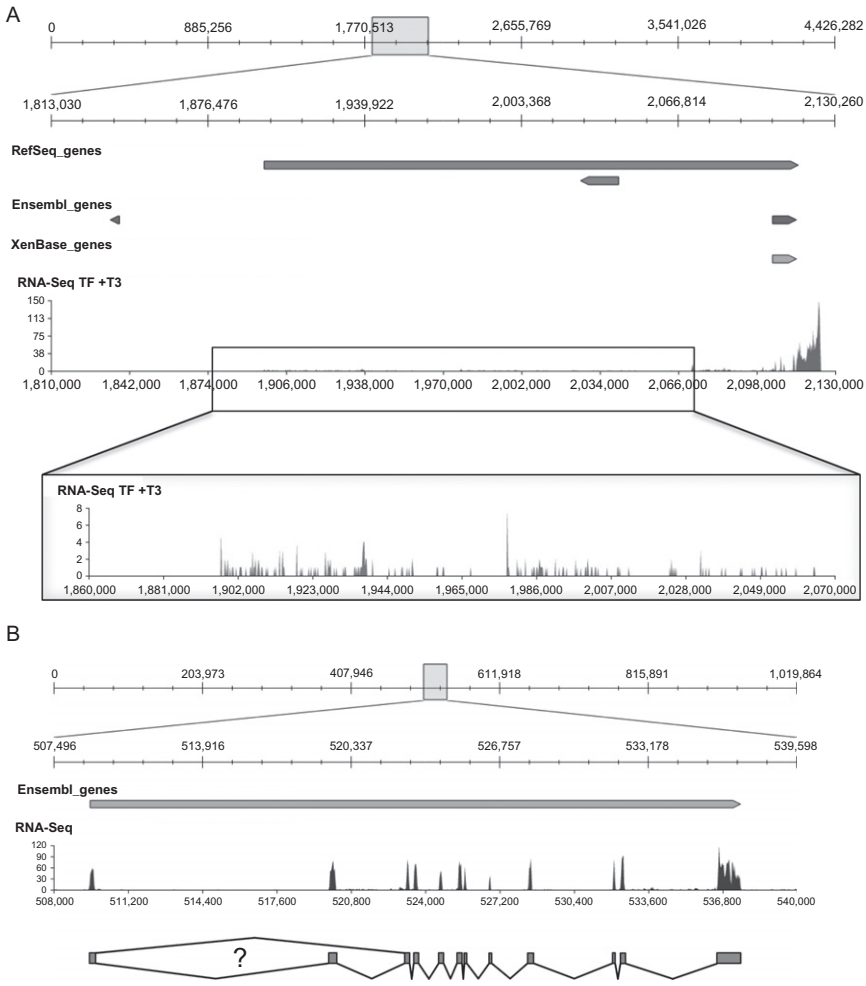

Figure 10.3 RNA-Seq profiling provides a quick overview of transcription units. (A) RNA-Seq profile at the TRβ locus. The first three tracks represent the gene models from the same databases as in Fig. 10.2. The fourth track represents the RNA-Seq profile obtained from the tail fin of premetamorphic tadpoles treated with T_3. The bottom part is a close-up of the TRβ transcription unit which is occulted by the very strong signal at the 3′ end of the gene. (B) Conventional RNA-Seq does not provide information on alternative splicing: Example of the wntless homolog (wls) on scaffold_431. The first track represents Ensembl gene models. The second track corresponds to the RNA-Seq profile. The third track corresponds to annotated exons and indicates two known alternative transcript are generated by exon skipping (connecting lines).

up to 210 kb upstream of the TSS annotated in XenBase or Ensembl databases. A limiting factor of conventional RNA-Seq is related to the length of the sequence reads (≤50 bp), that can be difficult to map when they overlap over two exons or when exons are small. Because of this, conventional RNA-Seq does not allow precise definition of exon boundaries and is not well suited for the definition of genes structure. This technical limitation is overcome with longer reads of 75–100 bp, and pair-end sequencing, which allows a better resolution when defining exons boundaries (Fig. 10.3B) (see below). This technology has been successfully used to refine current gene models in *X. tropicalis* (Akkers et al., 2009).

Several recent protocols allow longer sequence reads (≥100 bp) and pair-end sequencing, in which the two ends of each cDNA fragment are sequenced. Although the analysis is more complex, these protocols are well suited for *de novo* transcript reconstruction, exon definition, detection, and quantitative measure of alternative splicing. Two strategies can be used to process the sequenced reads (Garber et al., 2011): The reads can be assembled into transcripts, including splicing variants, which can be mapped to a reference genome, thus providing exon annotation. Alternatively, the reads can be mapped directly to the reference genome, without *de novo* assembly, and alternative splicing can be inferred from reads split in two and mapping over the ends of two distinct exons (also known as "spliced reads"). These methods provide a single base pair resolution of exon boundaries.

The 5′ end of transcripts are often missed and are poorly annotated. To this regard, the RNA-PET technology holds many promises. Originally developed to detect fusion transcripts in cancer cells, the RNA-PET specifically captures the 5′ and the 3′ ends of transcripts (Ruan & Ruan, 2011). The principle of this method is depicted in Fig. 10.4A. Full-length cDNA is synthesized from mRNA transcripts with CAP-Trapper and GsuI-dT(16) purification. A linker containing the binding site of the Mme1 (or *Eco*P15I) endonuclease is then ligated to the full-length cDNA, allowing it to be circularized. The circular DNA molecule is then restricted with *Mme*1, which cuts 27 bp downstream of its binding site, thus resulting in 27 bp tags of each end of the original transcript, spaced by the linker. Sequencing adaptors are then ligated for high-throughput sequencing. The resulting paired-end tags (PETs) are then mapped to the genome, thus providing experimental evidence for the 5′ and 3′ ends of transcripts, including potential alternative TSS and transcription termination sites (TTSs), but provides no information relative to the their internal structure (Fig. 10.4B). Thus, RNA-PET provides complementary information about the boundaries of the transcriptional

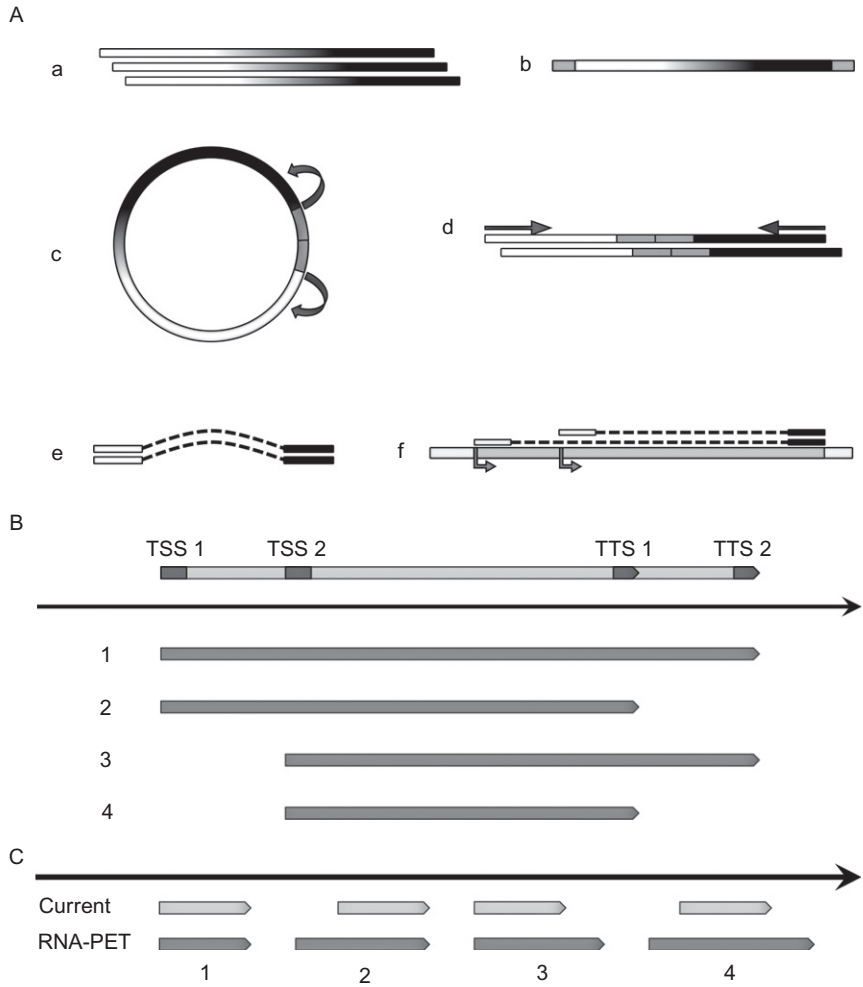

Figure 10.4 Principle of RNA-PET. (A) The starting point of RNA-PET is a library of full-length cDNAs made from poly-A RNA (a), on which linkers containing *Mme*1 (or *Eco*P15I) restriction sites are ligated (b). The resulting molecules are circularized and restricted with *Mme*1 (c), which cuts DNA 27 bp downstream of its binding site and thus produces a linear fragment containing the two ends of the original transcript separated by the linker (d) prior to deep sequencing (from each end). The resulting paired-end tags (PETs) contain information about the association of one transcript extremity to the other—respectively, white and black rectangles—(e). After mapping on the reference genome, this provides experimental evidence for the location of the 5′ and 3′ ends of transcripts (f). (B) Since RNA-PET captures the 5′ and 3′ ends of transcripts, it can be used to measure the relative abundance of transcripts that use alternative TSS or TTS (dark segments on the full gene model). (C) RNA-PET can improve the annotation of transcription units and thus gene boundaries. From left to right, RNA-PET determined gene boundaries (dark gray) can concord (1) with the current annotation (light gray), or extend in 5′ (2), 3′ (3), or both (4) extremities.

units, by confirming current models or redefining them at their 5′, 3′ (or both) ends (Fig. 10.4C). The potential of RNA-Seq coupled to RNA-PET in term of genome annotation is therefore tremendous.

3.2. NGS provides functional data to analyze TH-induced differential expression

Amphibian metamorphosis involves extensive modifications/remodeling of the organism as a whole. Since this biological process is strictly controlled by TH, documenting expression variation of TH-responsive genes is central to understand the molecular mechanisms controlling metamorphosis, decrypt the gene network controlled by TH, and its spatial and temporal dynamics. In the pre-NGS era, the simultaneous measure of the expression level of thousand of transcripts was carried out by microarrays. Yet, as discussed in a previous section, this high-throughput technology is not "naive" *per se*, in that the design of the microarrays is based on a prior knowledge of the transcripts to be detected. Indeed, RNA-Seq can benefit from the evolution and ongoing curation of a genome annotation, whereas detecting novel transcripts with microarrays would require a new design. In fact, important genes involved in amphibian metamorphosis might not even be represented on the chip (e.g., TRβ gene on *X. tropicalis* Affymetrix array, see above). Nowadays, in the NGS era, RNA-Seq represents a more naive, genome-wide alternative for measuring gene expression, in that the strength of the method is not based on a prior knowledge of transcripts, or gene annotation (Fig. 10.3). For example, in the case of TRβ, even if the gene is not well annotated, the RNA-Seq signal spreads over 210 kb and can be assigned to it *a posteriori*, whereas this is not possible with microarrays. RNA-Seq also has the benefit of having a much larger dynamic range than microarrays (Łabaj et al., 2011). In fact, the hybridization step in RNA-Seq is carried out *in silico*, which provides full control over it, and does not saturate. This is in contrast with microarrays, with which the hybridization signal can quickly reach the saturation limit of the photomultiplier. Once RNA-Seq data have been mapped to either a reference genome or transcriptome, several layers of analysis can be carried out. The absolute level of gene expression can be measured by computing the reads per kilobase of exon model per million of mapped reads (RPKM) value. Additionally, if several samples have been subject to the same analysis, one can perform differential analysis in a manner similar to that of microarray data, but using read count instead of photons. As shown in Fig. 10.5A, the expression of the TH regulated gene THb/ZIP is strongly induced upon T_3 treatment. Although the THb/ZIP gene

model and the Affymetrix microarray probesets partially overlap with the RNA-Seq signal (indicating that they can be used to assess the expression levels of the gene), a better measure of gene expression is achieved when counting the RNA-Seq reads along the entire gene. This is clearly shown in Fig. 10.5A, where the RNA-Seq signal extends up to 5 kb upstream of the annotated TSS for THb/ZIP, or even ~200 kb for TRβ (Figs. 10.3A and 10.5B). If read count was restricted to the annotated region, a lot of signal could actually be missed, potentially introducing biases comparable to those found in microarray analysis. In this context, using RNA-PET to define precise gene boundaries would prove particularly helpful in achieving better differential analysis of RNA-Seq data.

In addition, when using RNA-Seq derivatives (Wang et al., 2009), fragments spanning multiple exons can be easily mapped to the genome or transcriptome, giving precise data about the expression levels of the various alternative transcripts (Fig. 10.3B). Furthermore, beyond the strict definition of exon boundaries, RNA-PET can provide important data about TH effect on alternative TSS and TTS usage (Fig. 10.4B). The best-case scenario for accurately measuring gene expression with RNA-Seq requires high quality gene models.

4. TR INTERACTOME

4.1. Whole genome TR mapping

Recent advances provide the technologies to go beyond the simple measurement of transcripts abundance, making it possible to investigate the detailed molecular mechanisms involved in the transcriptional dynamics of TR mediated regulatory cascades.

Upon binding to T_3RE, TR recruits cofactors that will modify the transcriptional permissiveness of chromatin (Wong, Shi, & Wolffe, 1995). Although T_3RE are located near TH regulated genes (Ranjan et al., 1994), recent experimental evidence on the regulation of gene expression by transcription factors involves loop-mediated physical interactions between regulatory factors bound to DNA at distant enhancers and the core transcriptional machinery located at the TSS of regulated genes (Fullwood, Liu, et al., 2009). In fact, in a manner similar to other transcription factors, TR could regulate the expression of target genes over long genomic distances (up to several tens of kilobases, unpublished data), although some TR-binding site are located nearby TH-responsive genes.

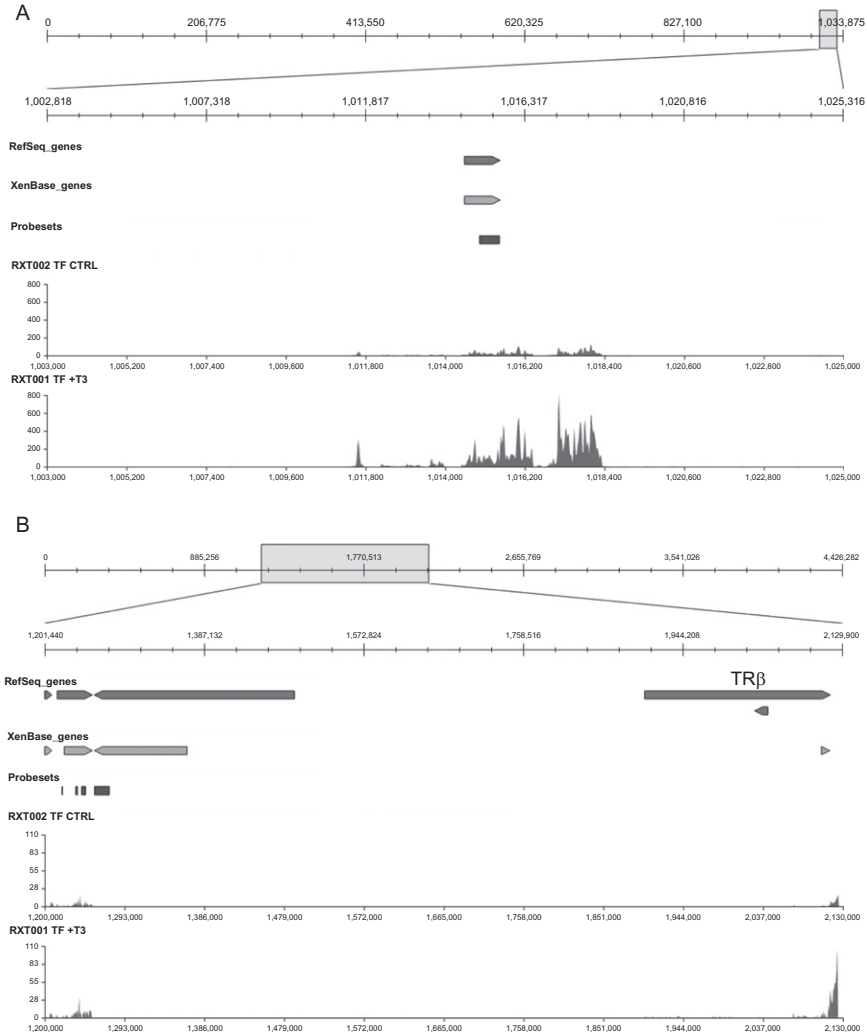

Figure 10.5 RNA-Seq is less limited than microarrays for quantitative measures of gene expression. (A) Genome view of the THb/ZIP locus. The first two tracks represent the RefSeq and XenBase gene models. The second track represents the (curated) probesets from the commercially available *X. tropicalis* Affymetrix microarray, with which the expression level of THb/ZIP appears induced 2.8 fold. The last two tracks correspond to the RNA-Seq profiles before and after T_3 treatment, respectively. In this case, the induction level is ~fivefolds, which fits better with RT-qPCR data. (B) Genome view of the TRβ locus. The tracks are the same as in A. No probeset is associated to the TRβ gene (on the right-hand side of the panel), as opposed to other genes (left-hand side). Thus, the Affymetrix microarray cannot monitor the expression of a key regulator of metamorphosis.

The canonical T_3RE sequence is well defined and it could be tempting to perform an exhaustive search of this motif genome wide with tools such as NHRscan (Sandelin & Wasserman, 2005). Unfortunately, given its small size and degeneracy, search for this motifs will be flooded by strong statistical noise (Fig. 10.6A). For example, more than 200,000 DR4 could be found across the genome. However, it is improbable that there are this many binding sites across the *X. tropicalis* genome, given that so far, the number of nuclear receptor binding site in a genome has been estimated between 2000 and 10,000 (Hamza et al., 2009; Nielsen et al., 2008; Welboren et al., 2009). In addition, identifying canonical DR4 motifs does not guaranty finding functional TR-binding sites. Indeed, TR was shown to bind other motifs (e.g., ER6 and DR1, see Desvergne, 1994). Nuclear receptors can also bind to motifs that depart significantly from the canonical motif (Zirngibl, Chan, & Aubin, 2008). Thus, it is imperative to use methods such as chromatin immunoprecipitation (ChIP) to provide experimental evidence of TR binding at specific genomic locations. To this end, protein–protein and protein–DNA interactions are stabilized by cross-linking, and chromatin is fragmented by sonication to a size range of 200–400 bp. The DNA fragments bound to the protein of interest are captured by immunoprecipitation with a specific antibody against the protein. After de-cross-linking and purification, the resulting DNA sample is enriched with the fragments originally associated with the protein. These samples can be subject to different detection techniques. With conventional ChIP–qPCR, the abundance of a known DNA fragment is measured by real-time quantitative PCR (qPCR). This method, which requires to know *a priori* the genomic region being monitored, has proved to be remarkably helpful in the context of amphibian metamorphosis to measure the differential binding patterns of TR in the vicinity of the TRβ gene, as well as the dynamics of histone modifications (Bilesimo et al., 2011). Although it is a very efficient technique, its throughput is very low and labor intense, since every location has to be tested individually by qPCR, and it requires to know *a priori* where to look for binding. With the ChIP-Seq technology, the detection method involves sequencing the purified DNA fragments. In this case, the number of sequence reads per genomic location provides a quantitative measure of DNA binding. This approach is naive and does not rely on prior knowledge of the putative binding sites, and rather provides genome-wide binding profiles (Fig. 10.6B).

The profiling of a few chromatin modification marks as well as RNA-Polymerase II occupancy has been carried out on *X. tropicalis* early

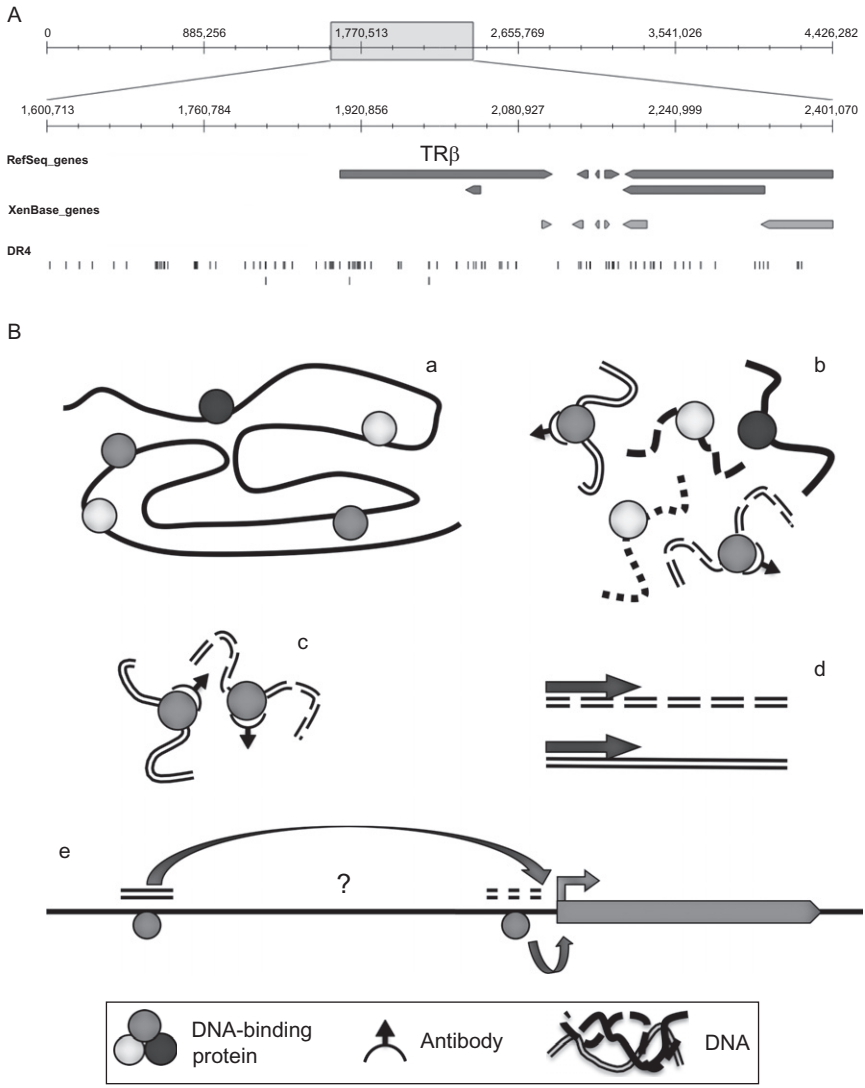

Figure 10.6 ChIP-Seq is a naive alternative to *in silico* predictions for mapping DNA-binding sites. (A) Genome view of the TRβ locus. The first two tracks represent the XenBase and Ensembl gene models, respectively. The third track represents 21,424 putative DR4 found in this region with a modified version of NHRscan (Sandelin & Wasserman, 2005). (B) Principle and limitations of ChIP-Seq. Physical interactions between chromatin components (protein, DNA) are stabilized by cross-linking (a). Chromatin is then fragmented by sonication in fragments of 200–400 bp (b). The DNA-binding sites are captured by immunoprecipitation with a protein-specific antibody (c). After cross-link reversal, the DNA sample is purified and sequenced (d). Binding density profiles are derived from sequenced reads after mapping to the reference genome, and provide an accurate description of the binding sites (e). Transcription factors (TFs) regulate the transcription of target genes after binding to DNA at enhancers, which can be located far from genes. Unfortunately, ChIP-Seq cannot connect DNA-binding sites of TFs to their target gene and thus cannot be used to describe the topology of gene regulatory networks.

development (Akkers et al., 2009), but not on amphibian metamorphosis. It would be of great interest to derive genome-wide TR-binding profiles to help understand the underlying mechanistic events of amphibian metamorphosis.

However, as seducing as it may sound, ChIP-Seq suffers from a major limitation as it provides no experimental evidence of the target gene(s) whose transcription is regulated by the DNA-bound transcription factor, if any. At best, one assumes that the target gene is located nearby the DNA-bound transcription factor, and infers such functional connection. Although this might be true for a few genes, this postulate is clearly not true genome wide (see below). It is clear that DNA-bound transcriptional regulators can act over large genomic regions and physically interact with the basal transcription machinery through DNA looping (Kim, Bresnick, & Bultman, 2009). Also, transcriptional regulators can interact with one another, thus adding more complexity (Fullwood, Liu, et al., 2009).

4.2. A need for 3D analysis

Recent technological breakthrough bypass the ChIP-Seq limitations and proceed a step further in the analysis of functional interactions between DNA-bound proteins. Ten years ago, the development of the chromosome conformation capture (3C) technology allowed to assess physical interactions between chromosome loci (Dekker, Rippe, Dekker, & Kleckner, 2002; de Wit & de Laat, 2012).The throughput was quite low in that 3C was used to detect the physical proximity of two DNA fragments in the nucleus in a "one to one" manner. It also required a prior knowledge of the genomic locations to probe. The first improvements of the 3C technology resulted in the circularized chromosome conformation capture (4C) technique, which allows to probe physical interactions in a "one to many" manner (Simonis et al., 2006; Zhao et al., 2006). In 2009, the Genome Institute of Singapore developed the chromatin interaction analysis by paired-end tags sequencing (ChIA-PET) (Fullwood, Liu, et al., 2009; Fullwood & Ruan, 2009), which combines the principles of ChIP-Seq and 4C, and produces two lines of evidence from a single experiment: a genome-wide binding profile together with a map of the physical interactions between protein-bound DNA regions. It is important to note that this method only reveals if two regions of DNA are close to each other but does not tell which or if any proteins is involved in the interaction. It does not either reveal any physical interactions among proteins, especially if no other proteins are analyzed. The

experimental procedure is quite complex and is aimed at capturing DNA-bound proteins engaging physical contacts with other protein-bound DNA complexes. The procedure can be summarized as follows (Fig. 10.7): DNA-bound protein complexes are captured by immunoprecipitation with a specific antibody, following chromatin fragmentation (this step is similar to that of a regular ChIP). However, as opposed to conventional ChIP, cross-linking is not reversed after this step. Instead, a double-stranded DNA linker containing the binding site of the restriction enzyme *Mme*1, which cuts DNA 27 bp downstream of its binding site is ligated to each end of the DNA fragment. The DNA fragments are then ligated. Two kinds of ligation events can take place: *cis*-ligation (self-ligation of a DNA fragment, providing ChIP-Seq like data but no information on long range interactions) and *trans*-ligation (ligation between two fragments far from each other on the genome but brought close to each other within the large protein–DNA complex). After ligation, the complexes are de-cross-linked, and the ligated DNA fragments are restricted with *Mme*1 to produce fragments containing pairs of ends ligated to the linker. After amplification and deep sequencing, they are mapped to the reference genome. One of the strengths of ChIA-PET relative to 3C and 4C is that it can probe "many to many" interactions of DNA regions associated with the protein, be it via interactions of the protein with other transcription factors (Fullwood, Liu, et al., 2009) or core components of the chromatin such as histones or other factors such as insulator-binding proteins (CCCTC-binding protein) (Chepelev, Wei, Wangsa, Tang, & Zhao, 2012; Handoko et al., 2011). As a consequence, in addition to profiling DNA-bound transcription factors, ChIA-PET also provides experimental evidence for physical interactions between DNA regions far away from each other, and thus can unambiguously identify direct target genes.

Using ChIA-PET on TR would help document with unprecedented details the topology and dynamics of regulatory networks of TH regulated genes. Importantly, this would not only provide a genome-wide map of TR binding but also identify direct TR target genes. Perhaps most of all, ChIA-PET will provide us with the opportunity to go beyond the local chromatin organization and probe higher orders of chromatin architecture. Our preliminary data show that indeed, TR can act over large genomic distances and that the local chromatin context might be marginally relevant to understand the mechanisms of TH-induced transcription cascades (unpublished data).

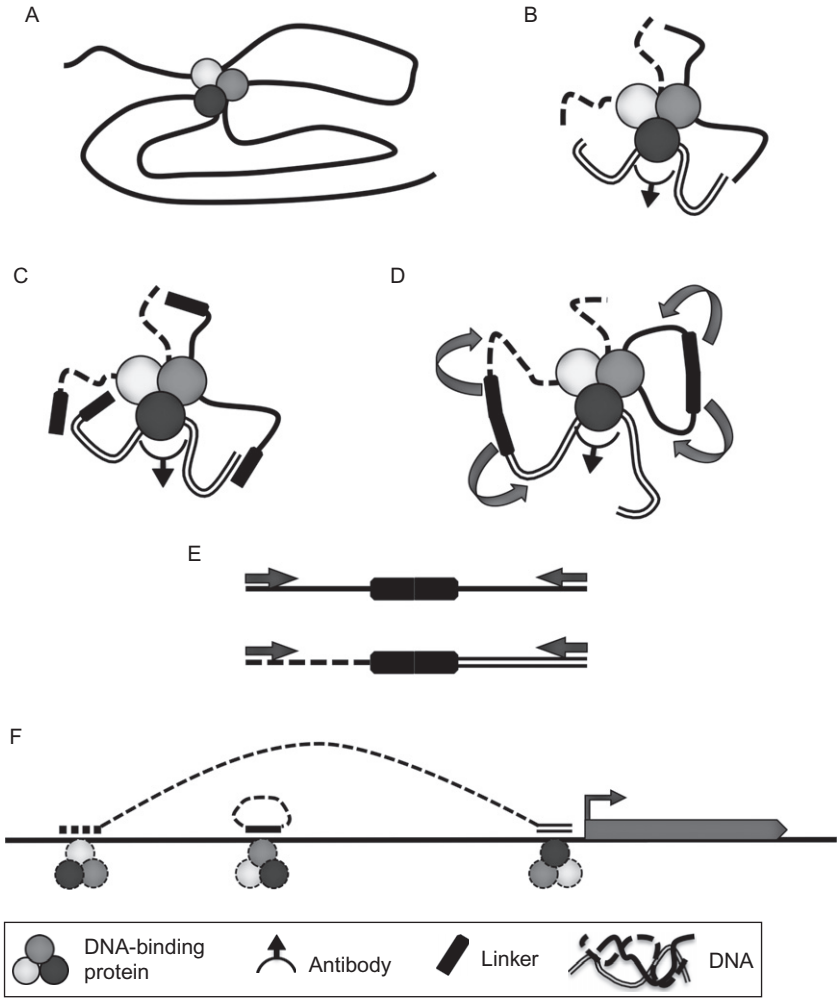

Figure 10.7 ChIA-PET describes DNA-binding sites and their physical interactions through DNA looping. The first two steps of ChIA-PET are similar to those of ChIP-Seq. Briefly, chromatin is cross-linked to stabilize the interactions between its proteins and DNA components and sonicated (A). Complexes are then purified with a specific antibody (B). Double stranded linkers containing the binding site of the *Mme*1 endonuclease are ligated to the ends of the fragmented DNA (C). Ligation can occur in *cis*, when a DNA fragment is circularized, or in *trans*, when two different DNA fragments are linked together (D). After restriction with *Mme*1, which cuts DNA 27 bp downstream of its binding site in the genomic fragment (D), the resulting linear fragments are sequenced, providing paired-end tags (PETs) reflecting the ligation products (E). After mapping to the reference genome, *cis*-ligation products are essentially similar to ChIP-Seq end products, and serve to derive binding profiles and identify DNA-binding sites. Given that the two tags of *trans*-ligation products map at distinct genomic locations, they do describe the network of physical interactions between the DNA-bound proteins (F). All the possible ligation products are not shown for the sake of simplicity.

5. CONCLUSION

Metamorphosis is a very common process which marks the transition between embryonic and adult life stages. Not only does it correspond to an ecological transition, in that the ecological niche often differs between the larva and the young adult, it also sets deep anatomical, physiological, cellular, and molecular changes. As such, metamorphosis is a deep, fundamental, and ancient biological process.

The problem of studying amphibian metamorphosis is that amphibian genomes are often polyploids, which adds extra levels of genetic redundancy and complexity, and limits to the tools which can be used to characterize the molecular mechanisms of metamorphosis. In this regard, the diploid *X. tropicalis* presents many advantages. Not surprisingly, at the beginning of genome sequencing, the scientific community quickly committed itself to sequence *X. tropicalis* genome and produced functional genomic resources. These were instrumental for the development of our current perception of the molecular mechanisms of *X. tropicalis* metamorphosis.

The recent breakthroughs in sequencing technologies push once again the scale and the precision of our molecular toolkits a step further. It is now possible to describe with exquisite details the genome-wide remodeling of transcriptional programs. Even better, such projects that used to require the combined workforce of one to several research institutes can now be carried out by a single lab. The promises of the molecular dissection of amphibian metamorphosis are enormous. The downside of it is that the current state of the *X. tropicalis* genome assembly and annotation does not reach the standard needed to make use of the full power of these technologies. It is thus important to revisit and update the *X. tropicalis* genomic resources.

The sequencing throughput has increased by three orders of magnitude in the past few years, posing many computing constraints: storage disks space is quickly saturated by the flood of sequencing data (because the advances of hard drive technologies is much slower than the progresses of sequencing throughput), large-scale analysis and processing often require large computing clusters (with their own constraints on dedicated facilities, power consumption, etc.), and data processing through large and complex bioinformatic pipeline. Theses weaknesses are counterbalanced (in part) by the progress made in software engineering with new and efficient data types, optimized database systems, and computing algorithms.

But perhaps the most challenging is our capacity, as scientists, to embrace the different layers of the technological fields deployed in such projects, which can go from breeding tadpoles in a tank to run metagenomic analysis on a distributed computing system. The success surely lies in the hands of next-generation scientists (Wickware, 2000).

ACKNOWLEDGMENTS

The work was supported by the "Centre National de la Recherche Scientifique," the "Muséum National d'Histoire Naturelle," "CRESCENDO," a European Integrated Project funding from FP6 (contract n°LSHM-CT-2005-018652), "IDEAL" a European Large Integrated Project funding from FP7 (n°259679), and "TRIGGER" (ANR-08-JCJC-0100-01). Part of this work was carried out in collaboration with the Genome Institute of Singapore, to whom we are grateful for introducing us to their technologies.

REFERENCES

Akkers, R. C., van Heeringen, S. J., Jacobi, U. G., Janssen-Megens, E. M., FranCoijs, K.-J., Stunnenberg, H. G., et al. (2009). A hierarchy of H3K4me3 and H3K27me3 acquisition in spatial gene regulation in Xenopus embryos. *Developmental Cell*, *17*, 425–434.

Altmann, C. R., Bell, E., Sczyrba, A., Pun, J., Bekiranov, S., Gaasterland, T., et al. (2001). Microarray-based analysis of early development in Xenopus laevis. *Developmental Biology*, *236*, 64–75.

Bilesimo, P., Jolivet, P., Alfama, G., Buisine, N., Le Mevel, S., Havis, E., et al. (2011). Specific histone lysine 4 methylation patterns define TR-binding capacity and differentiate direct T3 responses. *Molecular Endocrinology*, *25*, 225–237.

Birney, E. E., Clamp, M. M., & Durbin, R. R. (2004). GeneWise and genomewise. *Genome Research*, *14*, 988–995.

Bowes, J. B., Snyder, K. A., Segerdell, E., Gibb, R., Jarabek, C., Noumen, E., et al. (2008). Xenbase: A Xenopus biology and genomics resource. *Nucleic Acids Research*, *36*, D761–D767.

Brown, D. D., Wang, Z., Furlow, J. D., Kanamori, A., Schwartzman, R., Remo, B., et al. (1996). The thyroid hormone-induced tail resorption program during Xenopus laevis metamorphosis. *Proceedings of the National Academy of Sciences of the United States of America*, *93*, 1924–1929.

Buchholz, D. R., Heimeier, R. A., Das, B., Washington, T., & Shi, Y.-B. (2007). Pairing morphology with gene expression in thyroid hormone-induced intestinal remodeling and identification of a core set of TH-induced genes across tadpole tissues. *Developmental Biology*, *303*, 576–590.

Buckbinder, L., & Brown, D. D. (1992). Thyroid hormone-induced gene-expression changes in the developing frog limb. *The Journal of Biological Chemistry*, *267*, 25786–25791.

Buisine, N., & Sachs, L. (2009). Impact of genome assembly status on ChIP-Seq and ChIP-PET data mapping. *BMC Research Notes*, *2*, 257.

Cai, L., Das, B., & Brown, D. D. (2007). Changing a limb muscle growth program into a resorption program. *Developmental Biology*, *304*, 260–271.

Chepelev, I., Wei, G., Wangsa, D., Tang, Q., & Zhao, K. (2012). Characterization of genome-wide enhancer-promoter interactions reveals co-expression of interacting genes and modes of higher order chromatin organization. *Cell Research*, *22*, 490–503.

Das, B., Cai, L., Carter, M. G., Piao, Y.-L., Sharov, A. A., Ko, M. S. H., et al. (2006). Gene expression changes at metamorphosis induced by thyroid hormone in Xenopus laevis tadpoles. *Developmental Biology*, *291*, 342–355.

Das, B., Heimeier, R. A., Buchholz, D. R., & Shi, Y.-B. (2009). Identification of direct thyroid hormone response genes reveals the earliest gene regulation programs during frog metamorphosis. *The Journal of Biological Chemistry*, *284*, 34167–34178.

Dekker, J., Rippe, K., Dekker, M., & Kleckner, N. (2002). Capturing chromosome conformation. *Science*, *295*, 1306–1311.

de Wit, E., & de Laat, W. (2012). A decade of 3C technologies: Insights into nuclear organization. *Genes & Development*, *26*, 11–24.

Desvergne, B. (1994). How do thyroid hormone receptors bind to structurally diverse response elements? *Molecular and Cellular Endocrinology*, *100*, 125–131.

Eliceiri, B., & Brown, D. D. (1994). Quantitation of endogenous thyroid-hormone receptor-alpha and receptor-beta during embryogenesis and metamorphosis in Xenopus-laevis. *The Journal of Biological Chemistry*, *269*, 24459–24465.

Fairclough, L., & Tata, J. R. (1997). An immunocytochemical analysis of the expression of thyroid hormone receptor alpha and beta proteins during natural and thyroid hormone-induced metamorphosis in Xenopus. *Development, Growth & Differentiation*, *39*, 273–283.

Flamant, F., Gauthier, K., & Samarut, J. (2007). Thyroid hormones signaling is getting more complex: STORMs are coming. *Molecular Endocrinology*, *21*, 321–333.

Flicek, P., Amode, M. R., Barrell, D., Beal, K., Brent, S., Carvalho-Silva, D., et al. (2012). Ensembl 2012. *Nucleic Acids Research*, *40*, D84–D90.

Fullwood, M. J., & Ruan, Y. (2009). ChIP-based methods for the identification of long-range chromatin interactions. *Journal of Cellular Biochemistry*, *107*, 30–39.

Fullwood, M. J., Liu, M. H., Pan, Y. F., Liu, J., Xu, H., Mohamed, Y. B., et al. (2009). An oestrogen-receptor-alpha-bound human chromatin interactome. *Nature*, *462*, 58–64.

Fullwood, M. J., Wei, C.-L. C., Liu, E. T. E., & Ruan, Y. (2009). Next-generation DNA sequencing of paired-end tags (PET) for transcriptome and genome analyses. *Genome Research*, *19*, 521–532.

Furlow, J. D., & Brown, D. D. (1999). In vitro and in vivo analysis of the regulation of a transcription factor gene by thyroid hormone during Xenopus laevis metamorphosis. *Molecular Endocrinology*, *13*, 2076–2089.

Garber, M., Grabherr, M. G., Guttman, M., & Trapnell, C. (2011). Computational methods for transcriptome annotation and quantification using RNA-seq. *Nature Methods*, *8*, 469–477.

Gilchrist, M. J. (2012). From expression cloning to gene modeling: The development of Xenopus gene sequence resources. *Genesis*, *50*, 143–154.

Grimaldi, A., Buisine, N., Miller, T., Shi, Y.-B., & Sachs, L. M. (2012). Mechanisms of thyroid hormone receptor action during development: Lessons from amphibian studies. *Biochimica et Biophysica Acta*. http://dx.doi.org/10.1016/j.bbagen.2012.04.020.

Gronemeyer, H., & Laudet, V. (1995). Transcription factors 3: Nuclear receptors. *Protein Profile*, *2*, 1173–1308.

Hamza, M. S., Pott, S., Vega, V. B., Thomsen, J. S., Kandhadayar, G. S., Ng, P. W. P., et al. (2009). De-novo identification of PPARgamma/RXR binding sites and direct targets during adipogenesis. *PloS One*, *4*, e4907.

Handoko, L., Xu, H., Li, G., Ngan, C. Y., Chew, E., Schnapp, M., et al. (2011). CTCF-mediated functional chromatin interactome in pluripotent cells. *Nature Genetics*, *43*, 630–638.

Havis, E., Le Mevel, S., Morvan Dubois, G., Shi, D.-L., Scanlan, T. S., Demeneix, B. A., et al. (2006). Unliganded thyroid hormone receptor is essential for Xenopus laevis eye development. *The EMBO Journal*, *25*, 4943–4951.

Hawkins, R. D., Hon, G. C., & Ren, B. (2010). Next-generation genomics: An integrative approach. *Nature Reviews. Genetics*, *11*, 476–486.

Hellsten, U. U., Harland, R. M., Gilchrist, M. J., Hendrix, D., Jurka, J., Kapitonov, V., et al. (2010). The genome of the Western clawed frog Xenopus tropicalis. *Science*, *328*, 633–636.

Kawahara, A., Baker, B. S., & Tata, J. R. (1991). Developmental and regional expression of thyroid hormone receptor genes during Xenopus metamorphosis. *Development*, *112*, 933–943.

Kim, S.-I., Bresnick, E. H., & Bultman, S. J. (2009). BRG1 directly regulates nucleosome structure and chromatin looping of the alpha globin locus to activate transcription. *Nucleic Acids Research*, *37*, 6019–6027.

Klein, S., Strausberg, R., Wagner, L., Pontius, J., Clifton, S., & Richardson, P. (2002). Genetic and genomic tools for Xenopus research: The NIH Xenopus initiative. *Developmental Dynamics*, *225*, 384–391.

Łabaj, P. P., Leparc, G. G., Linggi, B. E., Markillie, L. M., Wiley, H. S., & Kreil, D. P. (2011). Characterization and improvement of RNA-Seq precision in quantitative transcript expression profiling. *Bioinformatics*, *27*, i383–i391.

Laudet, V. (2011). The origins and evolution of vertebrate metamorphosis. *Current Biology*, *21*, R726–R737.

Metzker, M. L. (2010). Sequencing technologies—The next generation. *Nature Reviews. Genetics*, *11*, 31–46.

Nagaraj, S. H., Gasser, R. B., & Ranganathan, S. (2007). A hitchhiker's guide to expressed sequence tag (EST) analysis. *Briefings in Bioinformatics*, *8*, 6–21.

Nielsen, R., Pedersen, T. A., Hagenbeek, D., Moulos, P., Siersbaek, R., Megens, E., et al. (2008). Genome-wide profiling of PPARgamma: RXR and RNA polymerase II occupancy reveals temporal activation of distinct metabolic pathways and changes in RXR dimer composition during adipogenesis. *Genes & Development*, *22*, 2953–2967.

Pruitt, K. D., Tatusova, T., Klimke, W., & Maglott, D. R. (2009). NCBI reference sequences: Current status, policy and new initiatives. *Nucleic Acids Research*, *37*, D32–D36.

Ranjan, M., Wong, J., & Shi, Y. B. (1994). Transcriptional repression of Xenopus TR beta gene is mediated by a thyroid hormone response element located near the start site. *Journal of Biological Chemistry*, *269*, 24699–24705.

Ruan, X., & Ruan, Y. (2011). Genome wide full-length transcript analysis using 5" and 3" paired-end-tag next generation sequencing (RNA-PET). *Methods in Molecular Biology*, *809*, 535–562.

Sachs, L., Damjanovski, S., Jones, P., Li, Q., Amano, T., Ueda, S., et al. (2000). Dual functions of thyroid hormone receptors during Xenopus development. *Comparative Biochemistry and Physiology. Part B, Biochemistry & Molecular Biology*, *126*, 199–211.

Sachs, L. M., & Shi, Y. B. (2000). Targeted chromatin binding and histone acetylation in vivo by thyroid hormone receptor during amphibian development. *Proceedings of the National Academy of Sciences of the United States of America*, *97*, 13138–13143.

Salamov, A. A., & Solovyev, V. V. (2000). Ab initio gene finding in Drosophila genomic DNA. *Genome Research*, *10*, 516–522.

Sandelin, A., & Wasserman, W. W. (2005). Prediction of nuclear hormone receptor response elements. *Molecular Endocrinology*, *19*, 595–606.

Santos, G. M. G., Fairall, L., & Schwabe, J. W. R. J. (2011). Negative regulation by nuclear receptors: A plethora of mechanisms. *Trends in Endocrinology and Metabolism*, *22*, 87–93.

Sboner, A., Mu, X. J., Greenbaum, D., Auerbach, R. K., & Gerstein, M. B. (2011). The real cost of sequencing: Higher than you think!. *Genome Biology*, *12*, 125.

Schiex, T., Moisan, A., & Rouzé, P. (2001). Lecture notes in computer science. In O. Gascuel & M.-F. Sagot (Eds.), Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg.

Shi, Y.-B. (1999). *Amphibian metamorphosis: From morphology to molecular biology* (1st ed). Wiley-Liss, New York.

Shi, Y. B., & Brown, D. D. (1993). The earliest changes in gene expression in tadpole intestine induced by thyroid hormone. *The Journal of Biological Chemistry*, *268*, 20312–20317.

Simonis, M., Klous, P., Splinter, E., Moshkin, Y., Willemsen, R., de Wit, E., et al. (2006). Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nature Genetics*, *38*, 1348–1354.

Tata, J. R. (1993). Gene expression during metamorphosis: An ideal model for post-embryonic development. *BioEssays*, *15*, 239–248.

Tymowska, J., & Fischberg, M. (1973). Chromosome complements of the genus Xenopus. *Chromosoma*, *44*, 335–342.

Voigt, J., Chen, J., Gilchrist, M., Amaya, E., & Papalopulu, N. (2005). Expression cloning screening of a unique and full-length set of cDNA clones is an efficient method for identifying genes involved in Xenopus neurogenesis. *Mechanisms of Development*, *122*, 289–306.

Wang, Z., & Brown, D. D. (1991). A gene-expression screen. *Proceedings of the National Academy of Sciences of the United States of America*, *88*, 11505–11509.

Wang, Z., & Brown, D. D. (1993). Thyroid hormone-induced gene expression program for amphibian tail resorption. *The Journal of Biological Chemistry*, *268*, 16270–16278.

Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: A revolutionary tool for transcriptomics. *Nature Reviews. Genetics*, *10*, 57–63.

Welboren, W.-J., van Driel, M. A., Janssen-Megens, E. M., van Heeringen, S. J., Sweep, F. C., Span, P. N., et al. (2009). ChIP-Seq of ERalpha and RNA polymerase II defines genes differentially responding to ligands. *The EMBO Journal*, *28*, 1418–1428.

Wickware, P. (2000). Next-generation biologists must straddle computation and biology. *Nature*, *404*, 683–684.

Wong, J., Shi, Y. B., & Wolffe, A. P. (1995). A role for nucleosome assembly in both silencing and activation of the Xenopus TR beta A gene by the thyroid hormone receptor. *Genes & Development*, *9*, 2696–2711.

Yaoita, Y., & Brown, D. D. (1990). A correlation of thyroid hormone receptor gene expression with amphibian metamorphosis. *Genes & Development*, *4*, 1917–1924.

Yaoita, Y., Shi, Y., & Brown, D. D. (1990). Xenopus laevis alpha and beta thyroid hormone receptors. *Proceedings of the National Academy of Sciences of the United States of America*, *87*, 7090–7094.

Zhao, Z., Tavoosidana, G., Sjölinder, M., Göndör, A., Mariano, P., Wang, S., et al. (2006). Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nature Genetics*, *38*, 1341–1347.

Zirngibl, R. A., Chan, J. S. M., & Aubin, J. E. (2008). Estrogen receptor-related receptor alpha (ERRalpha) regulates osteopontin expression through a non-canonical ERRalpha response element in a cell context-dependent manner. *Journal of Molecular Endocrinology*, *40*, 61–73.