

Published in final edited form as:

*Nat Methods*. 2013 December ; 10(12): . doi:10.1038/nmeth.2714.

## Assessment of transcript reconstruction methods for RNA-seq

Tamara Steijger<sup>1</sup>, Josep F Abril<sup>#2</sup>, Pär G Engström<sup>#1</sup>, Felix Kokocinski<sup>#3</sup>, RGASP Consortium<sup>4</sup>, Tim J Hubbard<sup>3</sup>, Roderic Guigó<sup>5,6</sup>, Jennifer Harrow<sup>3</sup>, and Paul Bertone<sup>1,7,8,9,\*</sup>

<sup>1</sup>European Molecular Biology Laboratory, European Bioinformatics Institute, Cambridge, UK

<sup>2</sup>Departament de Genètica, Facultat de Biologia, Universitat de Barcelona, Barcelona, Spain

<sup>3</sup>Wellcome Trust Sanger Institute, Cambridge, UK

<sup>5</sup>Center for Genomic Regulation, Barcelona, Spain

<sup>6</sup>Universitat Pompeu Fabra, Barcelona, Spain

<sup>7</sup>Genome Biology Unit, European Molecular Biology Laboratory, Heidelberg, Germany

<sup>8</sup>Developmental Biology Unit, European Molecular Biology Laboratory, Heidelberg, Germany

<sup>9</sup>Wellcome Trust – Medical Research Council Cambridge Stem Cell Institute, University of Cambridge, Cambridge, UK

\*Correspondence: bertone@ebi.ac.uk.

**Author contributions** JH, RG and TJH conceived and organised the study. Consortium members provided transcript models for evaluation. JH and PB coordinated the analysis, which was carried out by TS, JFA, PGE and FK. TS, PB and PGE wrote the manuscript with input from the other authors.

<sup>4</sup>**RGASP Consortium** Josep F Abril<sup>2</sup>, Martin Akerman<sup>11</sup>, Tyler Alioto<sup>12</sup>, Giovanna Ambrosini<sup>13,14</sup>, Stylianos E Antonarakis<sup>15</sup>, Jonas Behr<sup>16,17</sup>, Paul Bertone<sup>1,7,8,9</sup>, Regina Bohnert<sup>17</sup>, Philipp Bucher<sup>13,14,18</sup>, Nicole Cloonan<sup>19</sup>, Thomas Derrien<sup>5</sup>, Sarah Djebali<sup>6</sup>, Jiang Du<sup>20</sup>, Sandrine Dudoit<sup>21</sup>, Pär G Engström<sup>1</sup>, Mark Gerstein<sup>20,22,23</sup>, Thomas R Gingeras<sup>11</sup>, David Gonzalez<sup>5</sup>, Sean M Grimmond<sup>19</sup>, Roderic Guigó<sup>5,6</sup>, Lukas Habegger<sup>23</sup>, Jennifer Harrow<sup>3</sup>, Tim J Hubbard<sup>3</sup>, Christian Iseli<sup>18,24</sup>, Géraldine Jean<sup>17</sup>, André Kahles<sup>16,17</sup>, Felix Kokocinski<sup>3</sup>, Julien Lagarde<sup>5</sup>, Jing Leng<sup>23</sup>, Gregory Lefebvre<sup>13,18</sup>, Suzanna Lewis<sup>25</sup>, Ali Mortazavi<sup>26</sup>, Peter Niermann<sup>17</sup>, Gunnar Rätsch<sup>16,17</sup>, Alexandre Reymond<sup>27</sup>, Paolo Ribeca<sup>12</sup>, Hugues Richard<sup>28</sup>, Jacques Rougemont<sup>13,18</sup>, Joel Rozowsky<sup>22</sup>, Michael Sammeth<sup>5</sup>, Andrea Sboner<sup>22</sup>, Marcel H Schulz<sup>28</sup>, Steven MJ Searle<sup>3</sup>, Naryttza Diaz Solorzano<sup>18,24</sup>, Victor Solovvey<sup>29</sup>, Mario Stanke<sup>30</sup>, Tamara Steijger<sup>1</sup>, Brian Stevenson<sup>18,24</sup>, Heinz Stockinger<sup>18,24</sup>, Armand Valsesia<sup>18,24</sup>, David Weese<sup>31</sup>, Simon White<sup>3</sup>, Barbara J Wold<sup>32</sup>, Jie Wu<sup>11,33</sup>, Thomas D Wu<sup>34</sup>, Georg Zeller<sup>17</sup>, Daniel Zerbino<sup>1</sup>, Michael Q Zhang<sup>11</sup>

<sup>11</sup> Cold Spring Harbor Laboratory, New York, USA

<sup>12</sup> Centre Nacional d'Anàlisi Genòmica, Barcelona, Spain

<sup>13</sup> Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

<sup>14</sup> Swiss Institute for Experimental Cancer Research, Lausanne, Switzerland

<sup>15</sup> Department of Genetic Medicine and Development, University of Geneva Medical School, Geneva, Switzerland

<sup>16</sup> Computational Biology Center, Sloan-Kettering Institute, New York, USA

<sup>17</sup> Friedrich Miescher Laboratory of the Max Planck Society, Tübingen, Germany

<sup>18</sup> Swiss Institute of Bioinformatics, University of Lausanne, Lausanne, Switzerland

<sup>19</sup> Queensland Centre for Medical Genomics, The University of Queensland, St Lucia, Australia

<sup>20</sup> Department of Computer Science, Yale University, Connecticut, USA

<sup>21</sup> Division of Biostatistics, School of Public Health, University of California, Berkeley, California, USA

<sup>22</sup> Department of Molecular Biophysics and Biochemistry, Yale University, Connecticut, USA

<sup>23</sup> Program in Computational Biology and Bioinformatics, Yale University, Connecticut, USA

<sup>24</sup> Ludwig Institute for Cancer Research, Lausanne, Switzerland

<sup>25</sup> Genomics Division, Lawrence Berkeley National Laboratory, California, USA

<sup>26</sup> Department of Developmental and Cell Biology, University of California Irvine, California, USA

<sup>27</sup> Center for Integrative Genomics, University of Lausanne, Lausanne, Switzerland

<sup>28</sup> Max Planck Institute for Molecular Genetics, Berlin, Germany

<sup>29</sup> Department of Computer Science, Royal Holloway, University of London, London, UK

<sup>30</sup> Institute for Microbiology and Genetics, Göttingen, Germany

<sup>31</sup> Department of Mathematics and Computer Science, Freie Universität Berlin, Berlin, Germany

<sup>32</sup> Biology Division, California Institute of Technology, Pasadena, California, USA

<sup>33</sup> Department of Applied Mathematics and Statistics, Stony Brook University, New York, USA

<sup>34</sup> Bioinformatics and Computational Biology, Genentech, Inc., San Francisco, California, USA

**Software availability:** Source code for the evaluations performed in this study can be obtained from <https://github.com/RGASP-consortium>.

# These authors contributed equally to this work.

## Abstract

RNA sequencing (RNA-seq) is transforming genome biology, enabling comprehensive transcriptome profiling with unprecedented accuracy and detail. Due to technical limitations of current high-throughput sequencing platforms, transcript identity, structure and expression level must be inferred programmatically from partial sequence reads of fragmented gene products. We evaluated 24 protocol variants of 14 independent computational methods for exon identification, transcript reconstruction and expression level quantification from RNA-seq data. Our results show that most algorithms are able to identify discrete transcript components with high success rates, but that assembly of complete isoform structures poses a major challenge even when all constituent elements are identified. Expression level estimates also varied widely across methods, even when based on similar transcript models. Consequently, the complexity of higher eukaryotic genomes imposes severe limitations in transcript recall and splice product discrimination that are likely to remain limiting factors for the analysis of current-generation RNA-seq data.

## Introduction

High-throughput sequencing instruments necessitate a shotgun approach for all but the shortest target molecules. Full-length representation of most cellular RNAs from sequencing data requires computational reconstruction of transcript structures. The majority of such programs infer transcript models based on the accumulation of read alignments to the genome<sup>1-4</sup>, although some take the alternative approach of *de novo* reconstruction, where contiguous transcript sequences are assembled without use of a reference genome<sup>5-7</sup>.

Here we present a detailed evaluation of computational methods for transcript reconstruction and quantification from RNA-seq data, in a framework based on the ENCODE Genome Annotation Assessment Project (EGASP)<sup>8</sup>. Developers of leading software programs were invited to participate in a consortium effort, the RNA-seq Genome Annotation Assessment Project (RGASP), to benchmark methods to predict and quantify expressed transcripts from RNA-seq data. Results were evaluated from methods based on genome alignments (AUGUSTUS<sup>9</sup>, Cufflinks<sup>3</sup>, Exonerate<sup>10</sup>, GSTRUCT, iReckon<sup>2</sup>, mGene<sup>11</sup>, mTim, NextGeneid<sup>12</sup>, SLIDE<sup>4</sup>, Transomics, Trembly, Tromer<sup>13</sup>) as well as *de novo* assembly (Oases<sup>5</sup> and Velvet<sup>14</sup>). Our results identify aspects of RNA-seq analysis where current approaches are relatively adept, along with more challenging areas for future improvement.

## Results

A total of 24 transcript reconstruction protocols were evaluated, based on alternate parameter usage of 14 software packages (Supplementary Table 1 and Supplementary Note 1). Programs were run by the original developers, with the exception of Cufflinks, iReckon and SLIDE. To assess the ability of each method to interpret transcript expression from RNA-seq data without prior knowledge of gene content, programs were run without genome annotation, aside from iReckon and SLIDE which require such information. Performance was benchmarked relative to the subset of annotated exons to which RNA-seq reads mapped (coverage 1 read pair per 100 bp) and their corresponding transcripts (see Methods).

### Identification of annotated features

We first assessed the degree to which gene components reported by each algorithm matched the reference annotation at the nucleotide level. From the *C. elegans* data, the methods AUGUSTUS, mGene and Transomics displayed excellent performance in detecting exonic

bases, but also reported the expression of significant proportions of genomic sequence outside of reference exons (Fig. 1 and Supplementary Table 2). Recall (sensitivity) was generally lower for *D. melanogaster*, although most protocols exceeded 75% for both model organisms. Performance decreased for human data, where tradeoffs between precision and recall were more apparent. SLIDE and iReckon must be provided with gene annotation, and therefore outperformed most other methods. Even so, iReckon attained low precision at the nucleotide level, primarily due to the prediction of transcript isoforms with retained introns. AUGUSTUS, Exonerate, GSTRUCT, NextGeneid, Trembly and Velvet attained both precision and recall above 60% on the human data. The highest recall for methods without annotation was observed for Tromer and Cufflinks, albeit at the cost of low precision. These programs consistently displayed high sensitivity across the three species, but the precision rates for Tromer in particular indicates a tendency for overprediction. Notably, the AUGUSTUS *de novo* protocol, which predicts transcripts using the genomic sequence alone, reached nearly the same level of sensitivity as the corresponding protocol that also integrates RNA-seq data, but with significantly lower precision.

### Exon identification from RNA-seq data

We assessed the ability of each method to identify individual exons from RNAseq data relative to the reference annotation (Fig. 2). Accurate determination of transcription start and end sites is a known shortcoming of RNA-seq, and together with biological variation impairs the identification of transcript boundaries<sup>15-19</sup>. To mitigate this we allowed the 5'-ends of first exons and 3'-ends of terminal exons to differ from the reference coordinates (see Methods and Supplementary Fig. 2). Without these relaxed criteria, agreement in transcription start site and polyadenylation site positioning between predicted and annotated exons was extremely rare (Supplementary Fig. 3). Similarly, prediction accuracy for translation start and stop sites was low compared to internal exon boundaries, which can be inferred from spliced alignments (Supplementary Fig. 3 and Supplementary Table 3). Allowing for variable transcript boundaries led to significant improvements (Table 1, Supplementary Tables 4 and 5). Although most protocols exhibited the lowest precision for the human RNA-seq data, for all three species performance approached that of iReckon and SLIDE, despite the latter two benefiting from the use of high-quality gene annotation.

Coding exons can be identified directly from genomic sequence by the presence of translation start/stop sites and splice acceptors and donors. Programs such as AUGUSTUS, Exonerate, mGene, NextGeneid, Tromer and Transomics exploit these features to improve exon discovery. Of these, AUGUSTUS, mGene and Transomics identified a greater proportion of annotated coding exons than Exonerate, mTim, NextGeneid and Tromer (Supplementary Fig. 4). These methods augment data-driven transcript reconstruction with *ab initio* gene prediction, leading us to conclude that higher sensitivity measures are due to more extensive utilization of the underlying genomic sequence, thereby reducing the need for support from RNA-seq data.

We investigated the impact of sequencing depth on exon detection rates (Fig. 3a). Through the use of *ab initio* prediction, AUGUSTUS, mGene and Transomics were able to detect exons from protein-coding transcripts present at very low abundance. All other methods required a minimum average read depth to identify exons. Exon detection increased with sampling coverage at a roughly linear rate until reaching a plateau. One exception was Tromer, which often reported short exon fragments of 50–75 bp flanking introns without extending them to full exons (Supplementary Fig. 5). With increasing coverage, Tromer showed a tendency to predict very long exons spanning multiple annotated features. To a lesser extent Oases and Velvet also showed reduced performance for high-coverage exons (Supplementary Fig. 6).

The *ab initio* prediction advantage of AUGUSTUS, mGene and Transomics was lost for non-coding transcripts, which lack the consensus motifs exploited by these methods (Supplementary Fig. 7). Nevertheless, detection rates were similar to other protocols (Supplementary Fig. 8). Non-coding RNAs tend to be expressed at lower levels than protein-coding genes (Supplementary Fig. 9), and were detected with lower sensitivity even when controlling for differences in sequencing coverage (Fig. 3a and Supplementary Fig. 7). Exons of pseudogenes and processed transcripts were usually identified with higher frequency than those from long intergenic non-coding (linc) RNAs (Supplementary Fig. 10).

Several methods were unable to accurately determine the strand orientation for unspliced transcripts, resulting in low sensitivity for constituent exons (Fig. 2a and Supplementary Fig. 11). As the RNA-seq libraries were produced using standard protocols without strand specificity, directionality is inferred from sequence features such as translation and splicing signals. Accordingly, AUGUSTUS, mGene and NextGeneid outperformed other methods in the assignment of single-exon transcripts to the correct orientation (Fig. 2a and Supplementary Fig. 2). The overall impact on performance was limited to the proportion of single-exon transcripts in each species (3% for *C. elegans*, 14% for *D. melanogaster* and 7% for *H. sapiens*).

### Intron detection from RNA-seq data

The relative number and size of introns differ markedly between the three species used for this study (Supplementary Table 6). Overall AUGUSTUS, mGene and Transomics showed the highest intron detection rates (Fig. 3b). However, Transomics exhibited a sharper decline with increased intron length. This trend was apparent for all methods except Tromer, where a markedly lower detection rate was observed for introns shorter than 300 bp. To better characterize the differences in intron detection between methods, we classified reported introns based on overlap with known splice sites (Fig. 4). Most protocols predominantly detected known introns; several, however, also predicted a substantial number of introns with one or two novel splice sites. The highest frequencies of novel junctions were predicted by mGene, Transomics, Tromer, Velvet and the AUGUSTUS protocol that used only genomic sequence.

To explain this trend, we note that intron detection is highly dependent on the underlying read alignments, and that some aligners are more conservative than others (Engström et al., this issue). For example, PALMapper<sup>20</sup> was used as the alignment component in the mGene and mTim protocols. This aligner places more reads across unannotated splice sites than GEM<sup>21</sup>, GSNAP<sup>22</sup> and TopHat<sup>23,24</sup>; the latter programs form part of the NextGeneid, GSTRUCT and Cufflinks protocols, respectively.

### Assembly of exons into transcript isoforms

We next evaluated the performance of each method in linking exons into defined splice products. We initially determined the gene loci for which any expression was reported, regardless of whether a valid transcript was identified, followed by those consistent with at least one annotated isoform (Supplementary Fig. 12). Most algorithms detect transcription at over 80% of gene loci where expression is supported by RNA-seq reads. However, performance decreased substantially when considering genes for which at least one annotated transcript had been identified. For unguided transcript reconstruction, valid isoforms were assembled for roughly half of expressed genes on average (*H. sapiens* mean 41%, max 61%; *D. melanogaster* mean 55%, max 73%; *C. elegans* mean 50%, max 73%), and for those only one isoform was typically identified (Supplementary Fig. 13).

A significant reduction in sensitivity was also observed from the gene to transcript level, even when using the flexible evaluation mode for first and terminal exons (Fig. 5a and Supplementary Table 5). The best performing methods identified at most 56–59% of spliced protein-coding transcripts from *C. elegans* (AUGUSTUS, mGene and Transomics), 43% from *D. melanogaster* (AUGUSTUS) and merely 21% from *H. sapiens* (Trembly). Sensitivity increased by roughly 10% when partial isoform matches were considered, as did precision when including partial predictions consistent with annotated isoforms (Fig. 5a).

Greater sequencing depth improved transcript assembly for *D. melanogaster* and *C. elegans* (Supplementary Fig. 14a), whereas in *H. sapiens* transcript detection remained low despite sequencing coverage in excess of 4,000 read pairs/kb in exonic regions. Generally, at least one consistent isoform was identified for highly expressed genes: >50% in *D. melanogaster* and *C. elegans*, and >35% in *H. sapiens* (Supplementary Fig. 14b). Detection rates were even lower for non-coding RNAs (Supplementary Fig. 15). Pseudogenes were reported with similar frequency to protein-coding genes by AUGUSTUS, mGene, NextGeneid and Transomics, as pseudogenes retain partially intact coding sequences that can be identified by these methods (Supplementary Fig. 16).

The dramatic differences between species is further due to the tendency of methods to assign one splice product per gene (Supplementary Table 1). Whereas it is rare for genes in *C. elegans* and *D. melanogaster* to give rise to more than one or two transcript isoforms, human genes are annotated with an average of three to four, and it is unclear how many are simultaneously expressed. Assigning a single transcript model per gene will therefore impede the detection of multiple isoforms expressed in a given sample.

To identify the limiting factors in this process, for each method we calculated the number of known transcripts for which 1) all exons were identified, 2) exactly one exon was missing, 3) more than one exon was missing and 4) no exons were detected at all (Fig. 5b). The results clearly show that missing exons severely compromise transcript identification. For a significant fraction of transcripts not all exons are identified, ranging from 30% in *C. elegans* to greater than 60% in *H. sapiens*. Interestingly, while Trembly did not perform as well as AUGUSTUS, mGene and Transomics at the exon level, this method reported the highest number of transcripts for which all exons were represented from *H. sapiens* data. In contrast, AUGUSTUS, mGene and Transomics detected at least one exon for most transcripts. The remaining methods failed to identify any exons for nearly 20% of all transcripts expressed in the RNA-seq data. SLIDE exhibited the same trend despite the provision of annotated exon coordinates.

We then examined the topology of transcript structures to determine how well each method was able to link exons into complete isoforms. Even in cases where all exons of an annotated transcript had been identified, full isoforms were often not assembled (Supplementary Figure 16). For *C. elegans* and *D. melanogaster* most methods were able to reconstruct 60% of transcripts from the RNA-seq data. However, from the *H. sapiens* data less than 40% of known transcripts were assembled. Tromer stands out as an exception: the program identified all exons for relatively few genes, but once accounted for these were frequently linked into annotated transcript structures. Further inspection showed that these tended to be short isoforms comprising two to three exons on average, and thus represent a more tractable subset of the transcriptome.

Provision of transcript start and end sites gave iReckon an advantage for the more complex human transcriptome, as evidenced by increased accuracy in assembling partial transcripts. In contrast, SLIDE consults exon coordinates but ignores their connectivity, performing at a level similar to methods without any prior transcript-level information. Reported transcript

structures often differed substantially (Fig. 6a), and few were consistent across all methods (Supplementary Fig. 17). Pairwise agreement (Fig. 6b) was markedly higher for the model organisms than for human (median 25%), reflecting the number of partial isoforms identified as a function of transcriptome complexity.

### Quantification of expression levels from RNA-seq data

A common feature of transcript reconstruction software is the estimation of expression levels from transcribed genes. These are given as digital read counts normalized by transcript length and sequencing depth (reads per kilobase of exon model per million mapped reads, RPKM)<sup>25</sup>. RPKM values were reported at the transcript level from a subset of methods. A range of expression level distributions is evident (Supplementary Fig. 17), but generally show strong agreement between AUGUSTUS, iReckon, mGene and Trembly for all three RNA-seq data sets (Supplementary Fig. 18–21). One source of variation arises from gene loci where divergent or incomplete transcript models have been computed (Fig. 5a, Supplementary Figs. 22 and 23). However, expression level estimates can vary considerably even where concordant transcript structures are reported (Supplementary Fig. 17). Such differences were also apparent after scaling the RPKM distributions to equalize median expression values.

To establish independent expression level quantification we assayed a set of human genes using the NanoString nCounter amplification-free detection system<sup>26</sup> (Supplementary Table 7-9). Correlation between NanoString counts and RNA-seq RPKMs ranged from 0.35 for Transomics to 0.68 for Cufflinks (Fig. 6c and Supplementary Fig. 25). Many methods failed to report numerous targeted exons or junctions that were expressed according to NanoString counts. Read support at those loci was typically sparse, with 19 probes having no corresponding alignments from the RNA-seq data. These were, however, represented by low NanoString counts, indicating that the nCounter assay exhibits higher sensitivity for low-abundance transcripts than RNA-seq (Supplementary Fig. 25). For 10 of the unsupported NanoString probes, consistent isoforms were still reported by either AUGUSTUS, iReckon, mGene, SLIDE or Transomics. Thus, although the expression levels of these genes reflect the lower limits of detection for both technologies, sequencing reads dispersed over the gene body can allow for adequate transcript identification where *ab initio* methods or gene annotation were applied.

In general, all methods displayed higher identification rates for exons and junctions with higher NanoString counts, and reliable detection from RNA-seq data is dependent on read depth (Supplementary Figs. 26a and b). Nonetheless, each failed to report a subset of exons and junctions despite the availability of adequate RNA-seq alignments (Supplementary Fig. 26b). Comparing NanoString counts with RPKM values of the predominant isoform reported for each gene (irrespective of whether the targeted exon or junction was identified) improved correlation for most methods, and significantly for mTim and Transomics (Fig. 6c and Supplementary Fig. 28).

### Discussion

Technical limitations imposed by short-read sequencing lead to a number of computational challenges in transcript reconstruction and quantification. We carried out a communal effort to evaluate software tools from 14 developer teams, in realistic RNA-seq applications for species with high-quality reference genomes. Our assessments spanned a range of increasingly stringent metrics, from nucleotide-level correspondence with annotated transcript components to the assembly of full-length, multi-exonic splice variants.

AUGUSTUS, mGene and Transomics performed well at nucleotide- and exon-level detection for protein-coding genes. These methods predict transcript features from translation and splicing signals in the genomic sequence, reporting the model best supported by RNA-seq data. Combining *ab initio* prediction with experimental data can thus be more effective detecting genes expressed at low abundance, or from samples with low sequencing coverage. Even so, the benefits of this approach were most evident for *C. elegans* and lessened with increased transcriptome complexity.

The more challenging task of isoform reconstruction was reflected by lower performance in transcript-level evaluations. These results underscore the difficulty of transcript assembly, which relies on two outcomes: all exons comprising a given transcript must be identified, then connected to form the correct isoform structure. For most transcripts, automated methods failed to identify all constituent exons, and in cases where all exons were reported the protocols tested often failed to assemble them into complete isoforms. Whereas methods using *ab initio* prediction retain an advantage in detecting individual exons, others performed better at linking them together. For example, Trembly detected fewer exons than AUGUSTUS, mGene and Transomics, but identified the greatest number of valid transcript isoforms in *H. sapiens*. GSTRUCT did not reach the same detection rate as Trembly, but featured the highest overall precision of all methods. In *H. sapiens* and *C. elegans*, 57% and 60% of transcripts reported by GSTRUCT matched known isoforms and 76% in *D. melanogaster*.

AUGUSTUS, GSTRUCT, mGene, Transomics and Trembly thus outperformed other methods evaluated here, but no single protocol excelled at all metrics. Comparing the performance of AUGUSTUS with and without RNA-seq data as input revealed that using experimental evidence only slightly improved exon-level detection, but increased transcript-level precision. Transomics featured enhanced precision for high-abundance transcripts, but expression level differences had little impact on detection sensitivity. Precision was a consistent strength of GSTRUCT, whereas mGene exhibited diminished performance on human RNA-seq data underscoring that choice of method can depend on the organism under study.

Accurate identification of single-exon transcripts proved difficult for many of the methods tested. For spliced transcripts, the correct strand can typically be inferred from the genomic sequence at exon-intron boundaries. In the absence of this information single-exon transcripts were detected at much lower frequency. Antisense non-coding RNAs were another problematic class of transcripts missed by the methods evaluated. In these cases the detection of discrete transcription products may benefit from library construction methods that preserve strand orientation, such that exon coverage and read depth better reflect the expression of contiguous RNAs.

Significant variation was observed in the range of expression level estimates reported for transcripts arising from the same gene loci. This was exacerbated by non-uniform exon detection and linkage between methods, but was also apparent when similar or identical transcript structures were reported. Thus, it may be unreliable to directly compare gene-based RPKM values from sample data processed independently with different software tools. RNA-seq data to be compared from disparate sources should be treated in an identical manner from the initial processing steps. Where this is not possible, care should be taken to ensure that similar gene models have been identified, and RPKM distributions should be inspected before applying expression level thresholds in downstream analyses. Alternatively, uniform quantification of predicted transcripts can be performed with dedicated software<sup>27-29</sup>.

The potential for non-coding RNA discovery and characterization is a distinct advantage of RNA-seq over gene-based expression profiling. However this remains a challenging area for automated analysis methods. Performance is often impaired by lower expression levels of non-coding transcripts relative to many protein-coding genes, coupled with the inherent lack of translational features at the sequence level. The presence of open reading frames and translation start/stop signals allowed some methods to identify protein-coding transcripts even at very low expression levels, whereas the detection of non-coding RNAs at high confidence required much greater read depth. Sequencing coverage thus appears to be crucial for accurate non-coding RNA analysis.

Two of the methods evaluated here, Velvet and Oases, carry out direct transcriptome assembly rather than aligning sequencing reads to a reference genome. Velvet was originally developed for genome assembly from short read data. Oases is an adaption of that software, taking into account the unique challenges posed by transcriptome assembly from RNA-seq data, such as uneven exon representation due to amplification bias or tandem expression of multiple transcript isoforms. As expected, both methods attained similar performance at the nucleotide and exon levels. At the transcript level, however, Oases achieved higher detection rates for the more complex transcriptomes of *D. melanogaster* and *H. sapiens*. Although performance is generally enhanced by genome alignment strategies, studies based on organisms with a genome build of lower quality may benefit from a combined approach exploiting both genome alignments and *de novo* strategies.

The methods evaluated here can be applied to a range of analysis strategies, largely dependant on the state of the reference genome assembly and associated gene annotation for the target species (see Supplementary Table 10 and Supplementary Note 1). To improve the accuracy of existing annotation using RNA-seq data, both Cufflinks and iReckon consult known gene structures during the transcript assembly process and may be useful in refining the coordinates of exon and transcript boundaries. Where a finished genome and high-quality annotation are available, Cufflinks and rQuant (part of the mGene protocol) can be applied solely for transcript quantification, which can be further improved by correcting for fragment bias. Gene prediction algorithms such as AUGUSTUS and mGene can be used to automate the annotation of novel genomes, whereas RNA-seq experiments based on partial or low-quality genome builds can be approached with a *de novo* assembler like Oases. This last application is expected to receive increasingly wider attention with the continued sequencing of new genomes.

RNA-seq offers the potential to refine existing gene annotation through the discovery of novel exons and junction sites. However, unannotated transcript isoforms assembled from RNA-seq data should be interpreted with care, and those critical to an experimental study subjected to independent validation. The expression of multiple transcript isoforms and novel splice variants presents a major obstacle to accurate transcriptome reconstruction. Both exon identification and novel RNA discovery can improve with increased read depth, but the benefits of additional sampling to transcript assembly are inherently limited by the library construction requirements of current high-throughput sequencing platforms. Ultimately, the evolution of RNA-seq will move toward single-pass determination of intact transcripts. Third-generation instruments will realize that potential and inspire new computing approaches to meet the next wave of innovation in transcriptome analysis.

## Methods

### RNA-seq data

RNA-seq data were generated as part of the ENCODE<sup>30</sup> and modENCODE projects<sup>31</sup>, along with a third data set of compatible sequencing format and read depth, and represent



three widely-studied species: *Homo sapiens* (liver hepatocellular carcinoma cell line HepG2)<sup>32</sup>, *Drosophila melanogaster* (L3 stage larvae)<sup>33</sup>, and *Caenorhabditis elegans* (L3 stage larvae)<sup>34</sup>. These were chosen to reflect realistic examples of varying transcriptome complexity, and where high-quality annotated reference genomes are available. Libraries were prepared for the Illumina platform and sequenced in 76 nt paired-end format to obtain approximately 100 million read pairs per sample.

*H. sapiens* RNA-seq data correspond to ENCODE<sup>32</sup> HepG2 whole cell long polyA+ RNA CALTECH replicate 2, available from <http://www.encodeproject.org>. The *D. melanogaster* data set comprised a total of five sequencing runs from the modENCODE project<sup>33</sup> for three L3 stage larval samples, and can be obtained from the Sequence Read Archive (<http://www.ncbi.nlm.nih.gov/sra>) under accession numbers SRR023546, SRR023608, SRR023505, SRR027108 and SRR026433. The *C. elegans* RNA-seq data have previously been described<sup>34</sup> and are available under accession SRR065719. All of the data used in this study have been consolidated as a single experimental record in the ArrayExpress repository (<http://www.ebi.ac.uk/arrayexpress>) under accession E-MTAB-1730.

### Reference gene annotation

As not all genes are expressed in the samples used in the study, benchmarking methods against the entire set of annotated genes would underestimate transcript detection sensitivity. Therefore, we processed the genome annotations (*H. sapiens*: GENCODE<sup>30</sup> v15 (Ensembl release 70), *D. melanogaster*: FB2013\_01, *C. elegans*: WS200) to include only exons and transcripts with sufficient support in the RNA-seq data. Reads were mapped to the reference genomes using STAR version 2.2.0c, an independent RNA-seq aligner that is not a component in any of the evaluated transcript assembly methods<sup>35</sup>. To improve spliced alignment, STAR was provided with exon junction coordinates from the reference annotations. Default alignment parameters were used for the human data. For *D. melanogaster* and *C. elegans*, the intron size limit was reduced to 100,000 and 15,000 respectively (using options `--alignIntronMax` and `--alignMatesGapMax`). For each annotated exon, the read coverage (number of uniquely mapped read pairs divided by exon length) was computed and exons with a value below 0.01 fragments per bp were excluded from further analysis. Only transcripts for which all exons satisfied this criterion were included in transcript-level assessments. The threshold was determined by examining the exonic read coverage distribution, which consisted of three main features: a small peak at the low end (coverage < 0.01 fragments per bp), a dominant peak (coverage > 0.1) and a shoulder in between. Inspection of read alignments suggested that spurious reads are overrepresented in the minor peak, whereas the shoulder region comprises low-abundance transcripts and was therefore included in the analysis. To rule out potential bias imparted by the choice of alignment program, sensitivity and precision metrics were calculated for expressed genes based on several different spliced aligners (GSNAP, STAR and TopHat2) with no significant change to the results (Supplementary Fig. 28-30).

### Transcript prediction and assembly

Developer teams were provided with RNA-seq data and reference genome sequences for each species. To avoid potential biases, teams were not informed of the final evaluation criteria, and were not provided with gene annotation unless otherwise noted (e.g. iReckon, SLIDE). Developers providing transcript models for evaluation could not access submissions from other teams, and were prohibited from participating in the analysis phase as part of the study design. Details of transcript reconstruction protocols are provided in Supplementary Note 1.

## Data processing for Cufflinks, iReckon and SLIDE

RNA-seq reads were aligned with TopHat version 2.0.3 using parameters suited to each species. The genomes of *D. melanogaster* and *C. elegans* contain a high percentage of small introns (Supplementary Table 2); examining their size distributions led us to set the parameters `-i`, `--min-coverage-intron` and `--min-segment-intron` to 30 for *C. elegans*, 40 for *D. melanogaster* and 50 for *H. sapiens*.

Cufflinks was run with default settings except for the parameter `--min-intron-length` which was set to 30 for *C. elegans*, 40 for *D. melanogaster* and 50 for *H. sapiens*, consistent with the TopHat alignments. To maintain the greatest compatibility with submitted results that were computed without annotation, iReckon was run with the minimum annotation requirements, i.e. start and end sites of all annotated transcripts (not filtered by read coverage). SLIDE was run in discovery mode and provided with the full unfiltered annotation for each genome.

## Evaluation of prediction sets

Feature predictions were evaluated against the filtered reference annotation sets at four structural levels: nucleotide, exon, transcript and gene. The nucleotide-level metrics measure the ability of methods to identify exonic regions, ignoring the strand and exact boundaries of features. Nucleotide-level precision was computed as the number of genomic base pairs within both annotated and predicted exons, divided by the number of genomic base pairs within predicted exons. Similarly, nucleotide recall was computed as the number of genomic base pairs shared between annotated and predicted exons, divided by the number of genomic base pairs within annotated exons.

The exon-level metrics measure the ability of the different algorithms to identify the correct strand and boundaries of exons. Precision was calculated as the percentage of reported exons with an annotated counterpart, and recall denotes the percentage of annotated exons that were correctly assembled. Annotated exons were classified as first, internal, terminal and those comprising unspliced transcripts (single exons). Unless stated otherwise, a flexible evaluation mode was employed for first, terminal and single exons. Specifically, first and terminal exons were required to have correctly predicted internal borders only, and exons constituting unspliced transcripts were scored as correct if covered to at least 60% by a predicted transcript. Exons shared between different transcript isoforms were counted once. For comparison, certain analyses were also carried out using a fixed evaluation mode, where annotated and predicted exons were required to match exactly.

Transcript-level precision was computed as the percentage of reported spliced transcripts matching an annotated transcript, and recall as the percentage of annotated spliced transcripts with a counterpart in the transcript reconstruction output. Consistent with the flexible evaluation mode for exons (see above), transcript start and end sites were allowed to differ between reference and prediction, but splice sites were required to match exactly. Genes were scored as correctly predicted if at least one annotated transcript isoform in a given gene locus was correct. To estimate the degree of similarity between transcript predictions, a pairwise agreement score was calculated. The score  $a[i,j]$  denotes the fraction of transcription products predicted by protocol  $i$  consistent with those from protocol  $j$ . Methods were ordered by hierarchical clustering based on the distance metric  $1 - (a[i,j] + a[j,i]) / 2$ .

## Evaluation of transcript quantification

To compare transcript quantification results between methods, we identified for each annotated gene the corresponding predominant transcript reported; this was defined as the

transcript with the highest reported RPKM value among those isoforms intersecting annotated exons of the gene. A subset of human transcripts was quantified independently by NanoString assays. Genes of at least 1 kb in length, for which annotated exon-intron structures have been manually curated, and having at least two transcripts satisfying these criteria were selected. A total of 109 genes were targeted by 141 distinct probes, designed against specific exons or splice junctions.

NanoString counts were compared to the highest RPKM value reported for transcript isoforms consistent with the probe design (correlation  $r_c$ ) or for any isoform from the locus (correlation  $r_l$ ). Predicted transcripts were required to contain the exon or junction targeted by the NanoString probe. Where multiple such transcripts were reported for the same gene, the highest RPKM value was used. Where no such transcript was reported, an RPKM of zero was assigned. Percentages reflect the probes for which transcripts satisfying these criteria were reported. Pearson's  $r$  was calculated based on the log-transformed NanoString counts and RNA-seq RPKM values. Expression values were incremented by 1 prior to transformation to avoid infinite numbers.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

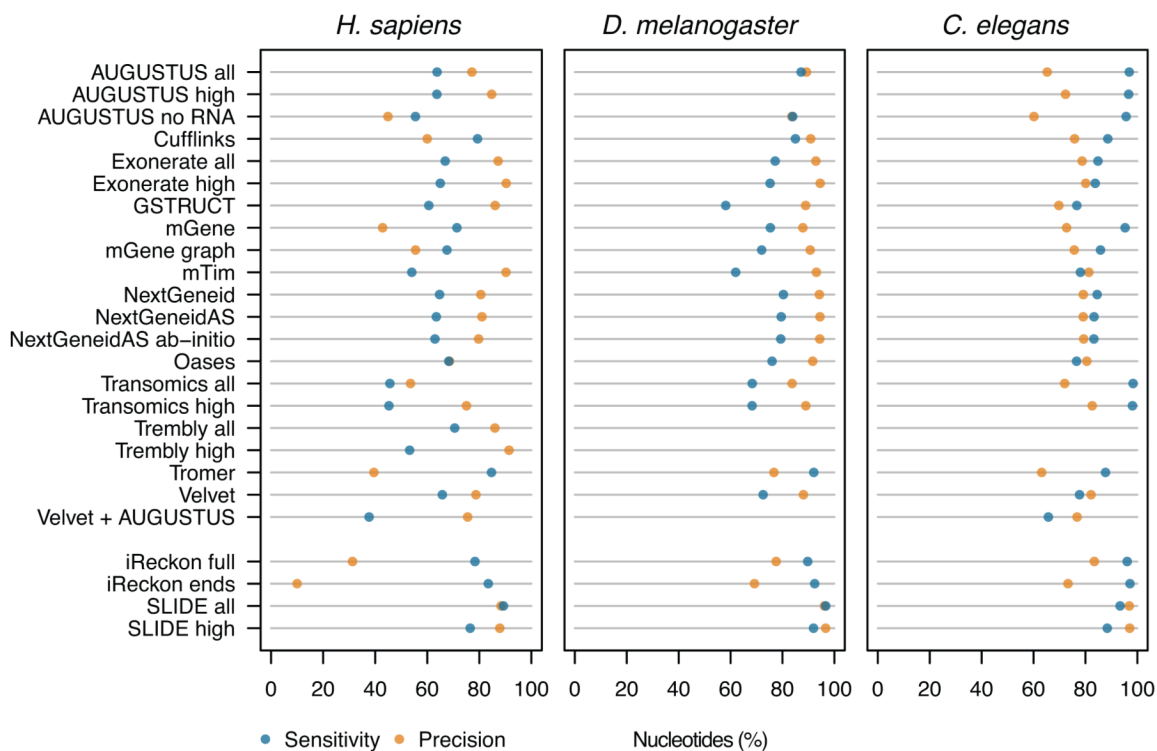
## Acknowledgments

This work was supported by EMBL, NIH/NHGRI grants U54HG004555 and U54HG004557, Wellcome Trust grant WT098051, and grants BIO2011-26205 and CSD2007-00050 from the Ministerio de Educación y Ciencia.

## References

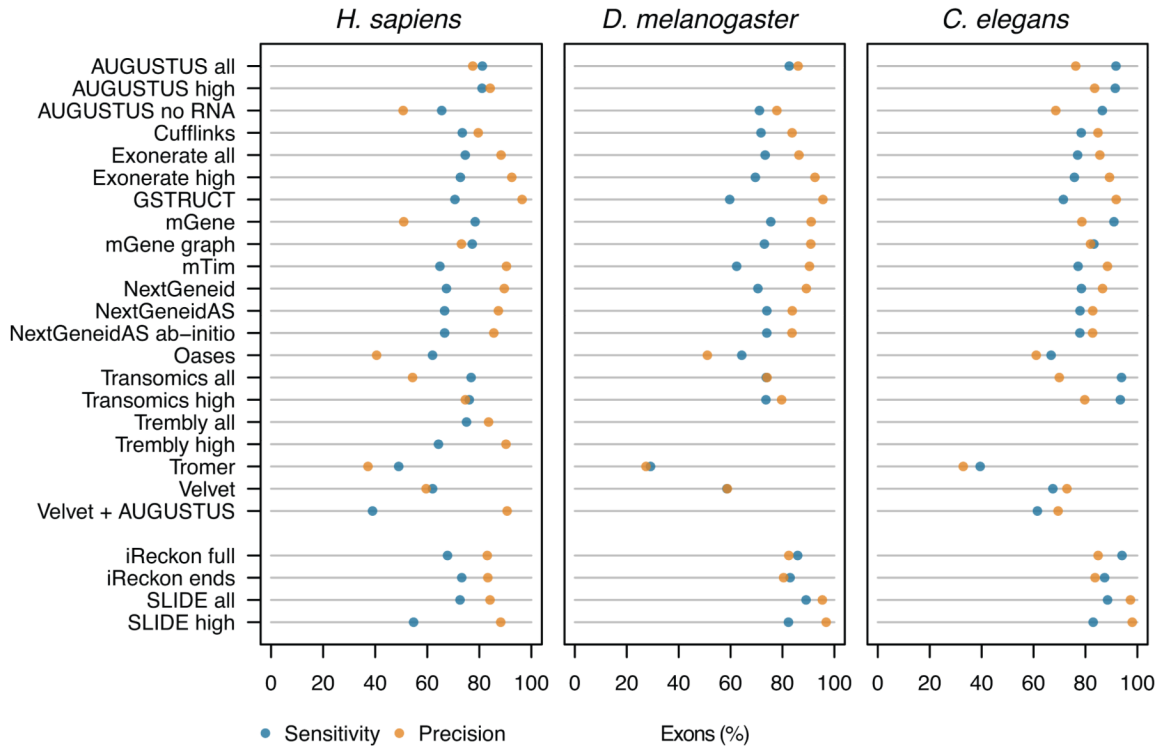
1. Trapnell C, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* 2010; 28:511–515. [PubMed: 20436464]
2. Mezlini AM, et al. iReckon: simultaneous isoform discovery and abundance estimation from RNA-seq data. *Genome Res.* 2013; 23:519–529. [PubMed: 23204306]
3. Roberts A, Pimentel H, Trapnell C, Pachter L. Identification of novel transcripts in annotated genomes using RNA-Seq. *Bioinformatics.* 2011; 27:2325–2329. [PubMed: 21697122]
4. Li JJ, Jiang C-R, Brown JB, Huang H, Bickel PJ. Sparse linear modeling of next-generation mRNA sequencing (RNA-Seq) data for isoform discovery and abundance estimation. *Proc. Natl. Acad. Sci. U.S.A.* 2011; 108:19867–19872. [PubMed: 22135461]
5. Schulz MH, Zerbino DR, Vingron M, Birney E. Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics.* 2012; 28:1086–1092. [PubMed: 22368243]
6. Grabherr MG, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 2011; 29:644–652. [PubMed: 21572440]
7. Robertson G, et al. De novo assembly and analysis of RNA-seq data. *Nat. Methods.* 2010; 7:909–912. [PubMed: 20935650]
8. Guigo R, et al. EGASP: the human ENCODE Genome Annotation Assessment Project. *Genome Biol.* 2006; 7(Suppl 1):1–31.
9. Stanke M, et al. AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res.* 2006; 34:W435–9. [PubMed: 16845043]
10. Slater GSC, Birney E. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics.* 2005; 6:31. [PubMed: 15713233]
11. Schweikert G, et al. mGene: accurate SVM-based gene finding with an application to nematode genomes. *Genome Res.* 2009; 19:2133–2143. [PubMed: 19564452]
12. Blanco E, Parra G, Guigo R. Using geneid to identify genes. *Curr Protoc Bioinformatics.* 2007; Chapter 4(Unit 4.3):1–28.

13. Sperisen P, et al. trome, trEST and trGEN: databases of predicted protein sequences. *Nucleic Acids Res.* 2004; 32:D509–11. [PubMed: 14681469]
14. Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* 2008; 18:821–829. [PubMed: 18349386]
15. Lenhard B, Sandelin A, Carninci P. Metazoan promoters: emerging characteristics and insights into transcriptional regulation. *Nat. Rev. Genet.* 2012; 13:233–245. [PubMed: 22392219]
16. Di Giammartino DC, Nishida K, Manley JL. Mechanisms and consequences of alternative polyadenylation. *Mol. Cell.* 2011; 43:853–866. [PubMed: 21925375]
17. Tian B, Hu J, Zhang H, Lutz CS. A large-scale analysis of mRNA polyadenylation of human and mouse genes. *Nucleic Acids Res.* 2005; 33:201–212. [PubMed: 15647503]
18. Batut P, Dobin A, Plessy C, Carninci P, Gingeras TR. High-fidelity promoter profiling reveals widespread alternative promoter usage and transposon-driven developmental gene expression. *Genome Res.* 2013; 23:169–180. [PubMed: 22936248]
19. Shepard PJ, et al. Complex and dynamic landscape of RNA polyadenylation revealed by PAS-Seq. *RNA.* 2011; 17:761–772. [PubMed: 21343387]
20. Jean G, Kahles A, Sreedharan VT, De Bona F, Rätsch G. RNA-Seq read alignments with PALMapper. *Curr Protoc Bioinformatics.* 2010; Chapter 11(Unit 11.6):1–37.
21. Marco-Sola S, Sammeth M, Guigo R, Ribeca P. The GEM mapper: fast, accurate and versatile alignment by filtration. *Nat. Methods.* 2012; 9:1185–1188. [PubMed: 23103880]
22. Wu TD, Nacu S. Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics.* 2010; 26:873–881. [PubMed: 20147302]
23. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics.* 2009; 25:1105–1111. [PubMed: 19289445]
24. Kim D, et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* 2013; 14:R36. [PubMed: 23618408]
25. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods.* 2008; 5:621–628. [PubMed: 18516045]
26. Kulkarni MM. Digital multiplexed gene expression analysis using the NanoString nCounter system. *Curr Protoc Mol Biol.* 2011 Chapter 25, Unit25B.10.
27. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics.* 2011; 12:323. [PubMed: 21816040]
28. Katz Y, Wang ET, Airolidi EM, Burge CB. Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nat. Methods.* 2010; 7:1009–1015. [PubMed: 21057496]
29. Bohnert R, Rätsch G. rQuant.web: a tool for RNA-Seq-based transcript quantitation. *Nucleic Acids Res.* 2010; 38:W348–51. [PubMed: 20551130]
30. Harrow J, et al. GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res.* 2012; 22:1760–1774. [PubMed: 22955987]
31. modENCODE Consortium. et al. Identification of functional elements and regulatory circuits by *Drosophila* modENCODE. *Science.* 2010; 330:1787–1797. [PubMed: 21177974]
32. Djebali S, et al. Landscape of transcription in human cells. *Nature.* 2012; 489:101–108. [PubMed: 22955620]
33. Graveley BR, et al. The developmental transcriptome of *Drosophila melanogaster*. *Nature.* 2011; 471:473–479. [PubMed: 21179090]
34. Mortazavi A, et al. Scaffolding a *Caenorhabditis* nematode genome with RNA-seq. *Genome Res.* 2010; 20:1740–1747. [PubMed: 20980554]
35. Dobin A, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics.* 2013; 29:15–21. [PubMed: 23104886]



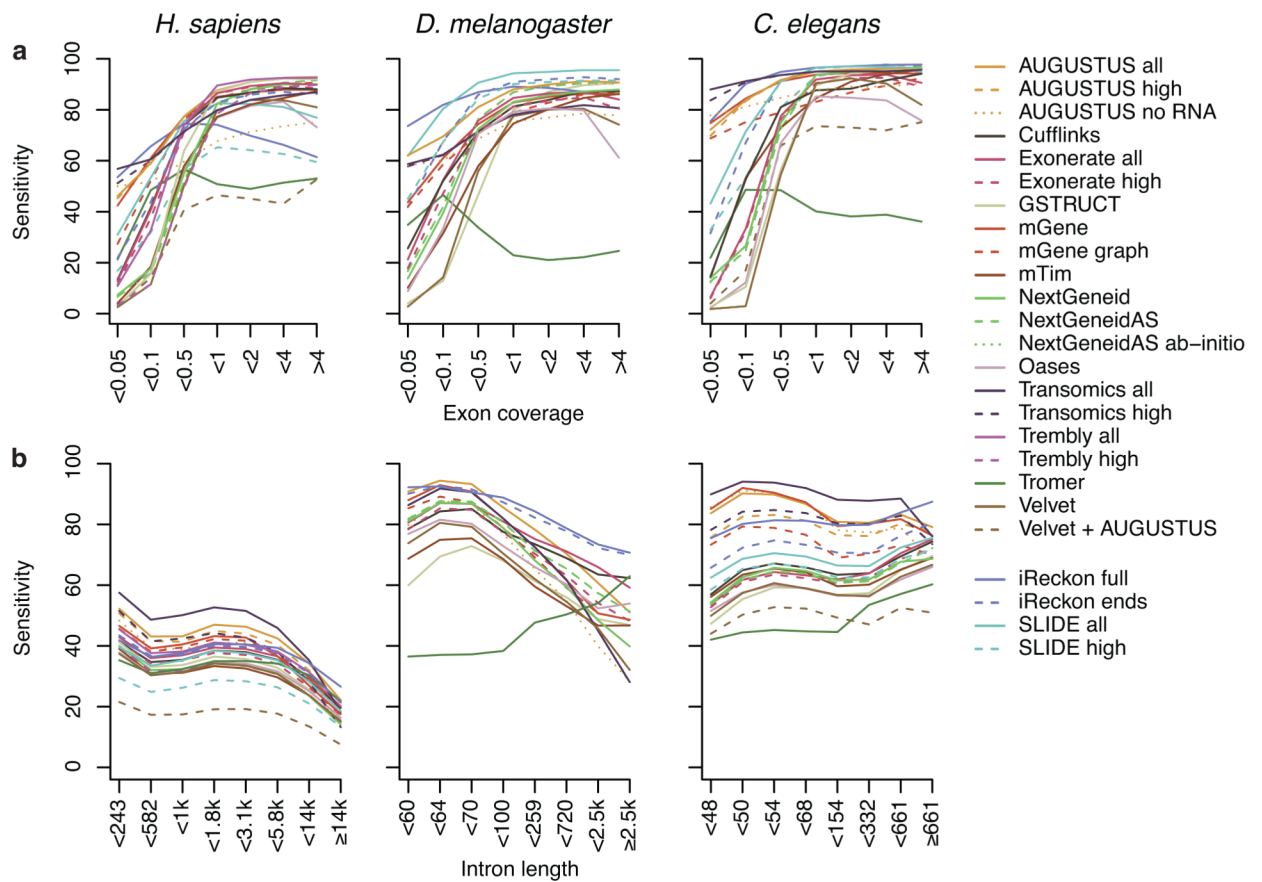
**Figure 1. Summary of nucleotide- and exon-level performance for the methods evaluated.**

(a) Performance at detecting exonic nucleotides. Recall (green) indicates the proportion of known exon sequence in each genome covered by assembled transcripts, and precision (red) the proportion of reported expressed sequence confined to known exons. (b) Performance at detecting individual exons, shown as the percentage of reference exons with a matching feature in the submission (recall, green), and the proportion of reported exons that agree with annotation (precision, red). Programs run with gene annotation are grouped separately.



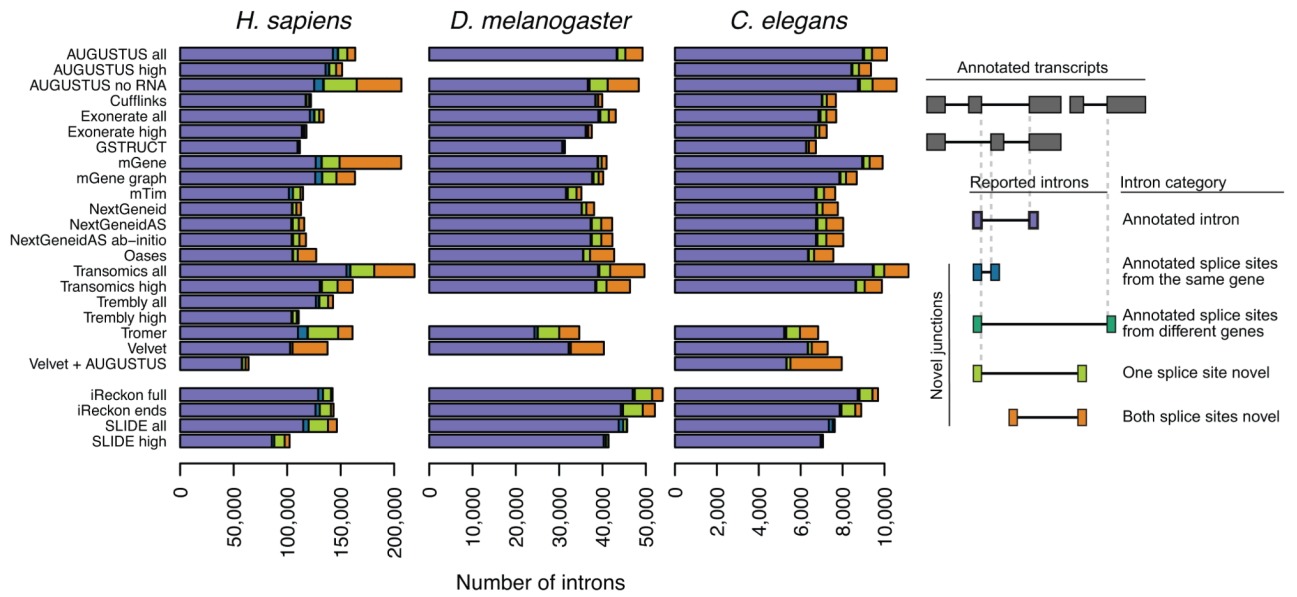
**Figure 2. Influence of exon rank and sequencing coverage on detection performance.**

(a) Detection sensitivity for annotated human exons classified as first, internal, terminal or single (i.e., those comprising an entire transcript). Exon boundaries were required to be predicted exactly as annotated (left, center) or by employing relaxed criteria for exons at transcript boundaries (right). See Supplementary Figure 3 for a corresponding analysis for the model organisms. (b) Sensitivity for detection of annotated exons stratified by read depth.



**Figure 3. Intron detection performance.**

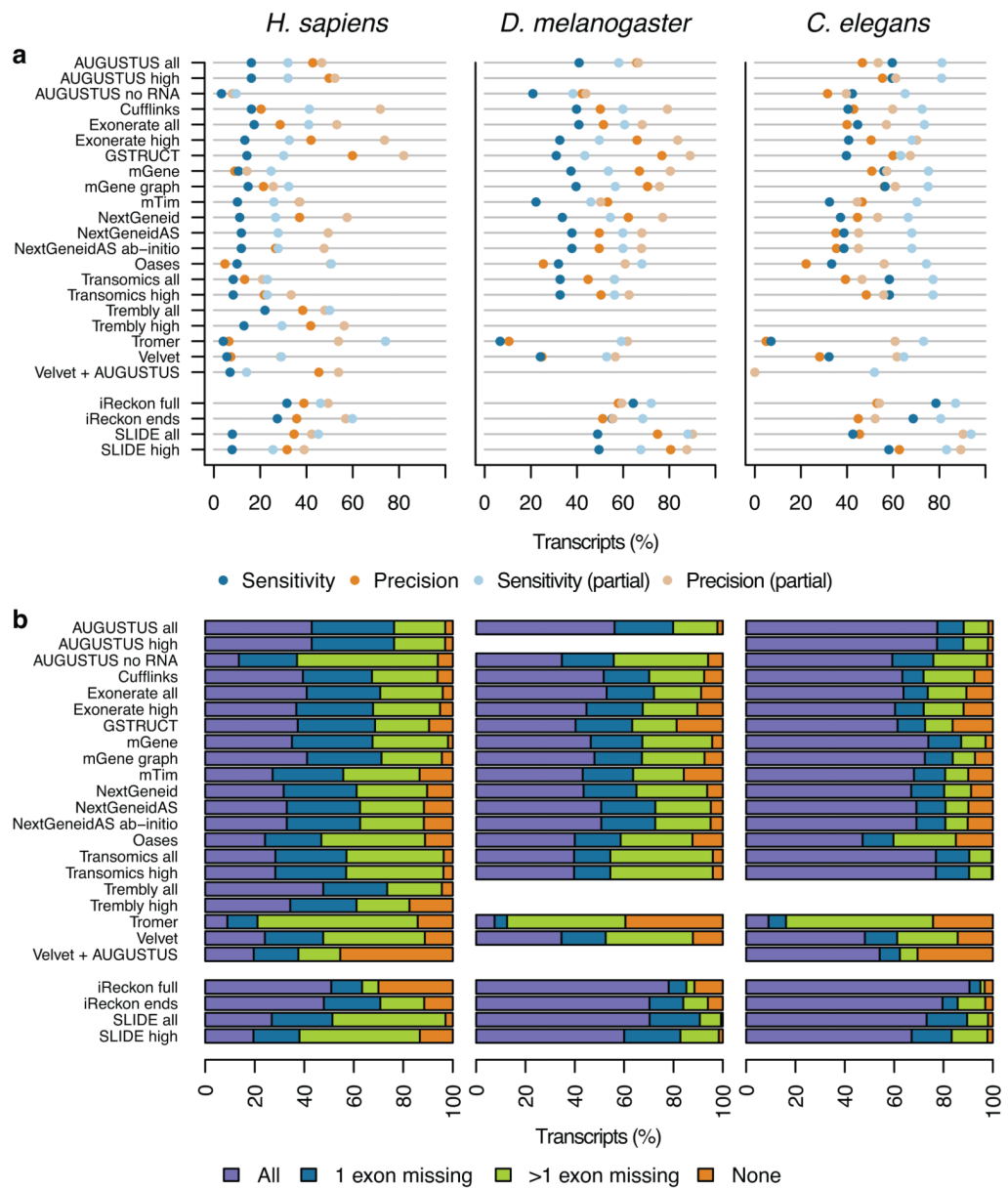
(a) Annotated introns were binned on length and sensitivity was calculated separately for each bin. (b) Reported introns were classified by overlap with splice sites annotated in the reference gene sets.



**Figure 4. Transcript assembly performance.**

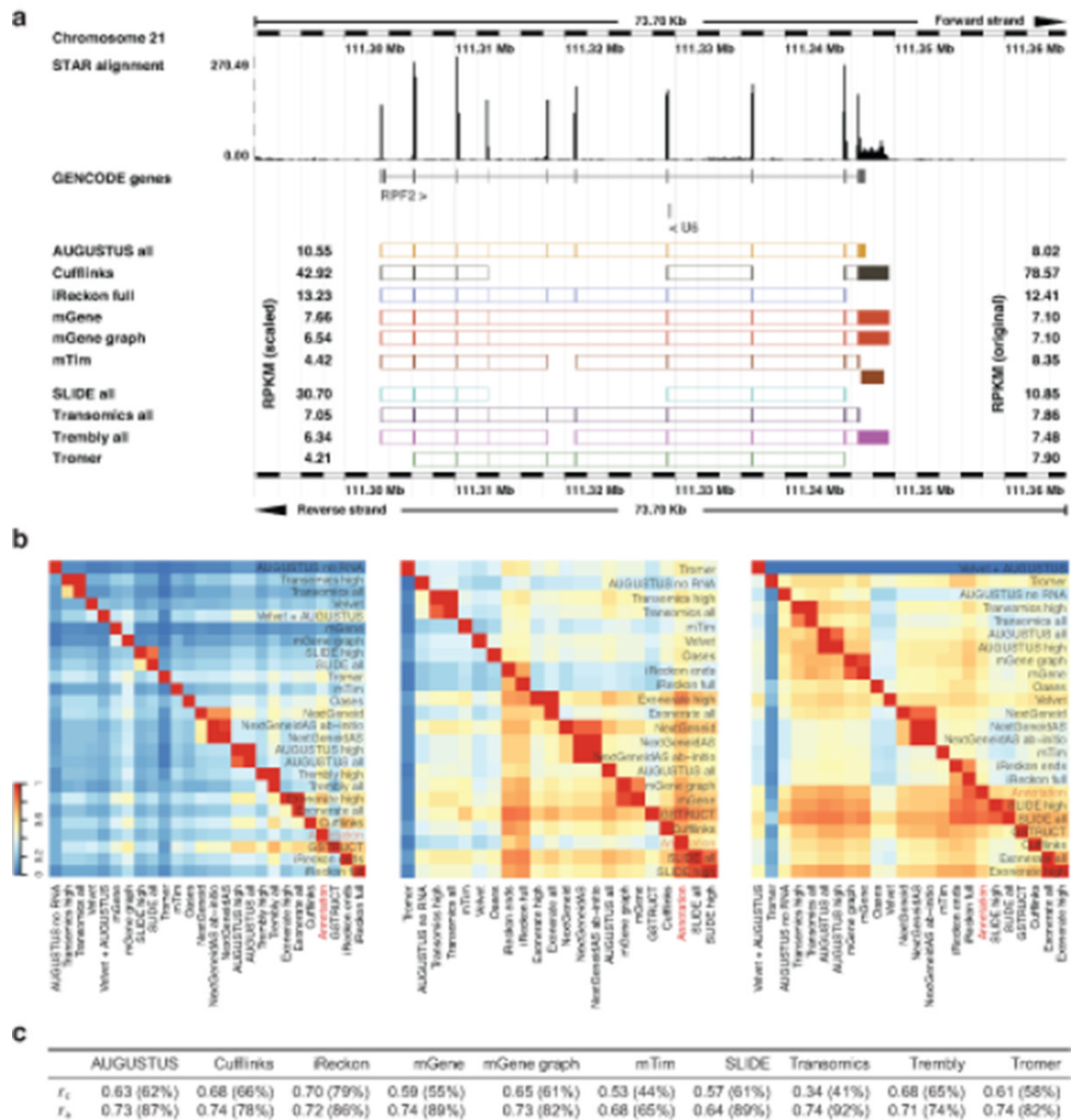
(a) Reference transcripts with a matching submission entry (transcript-level recall, green) and reported transcripts that match the reference (transcript-level precision, red). (b) Transcripts for which various subsets of constituent exons have been reported. (c) Percentage of transcripts, for which all exons have been identified, that were correctly assembled to a full-length annotated splice variant.





**Figure 5. Transcript assembly performance.**

(a) Reference transcripts with a matching submission entry (transcript-level sensitivity, blue) and reported transcripts that match the reference (transcript-level precision, orange). (b) Transcripts for which various subsets of constituent exons have been reported.



**Figure 6. Examples of transcript calls and expression level estimates.**

(a) The upper tracks show RNA-seq read coverage (from STAR alignments; see Methods) and annotated genes. Exon predictions from the 10 methods that quantified transcripts are illustrated below the annotated gene by colored boxes. Exons predicted to belong to the same transcript isoform are connected. Original and median-scaled RPKM values are presented to the right and left, respectively, of the transcript models. For the gene *RPF2*, all methods reported different isoforms and expression levels. Where multiple overlapping isoforms were reported, that with the higher RPKM was chosen for visualization, and spliced isoforms were prioritized over unspliced ones. The noncoding RNA *U6* is not expressed. (b) Heat maps illustrate pairwise agreement between reported transcript isoforms for *H. sapiens* (left), *D. melanogaster* (center) and *C. elegans* (right). (c) Correlation between reported RPKM values and NanoString counts (Pearson  $r$  of log-transformed values). NanoString counts were compared to the highest RPKM value reported for

transcript isoforms consistent with the probe design (correlation  $r_c$ ) or for any isoform from the locus (correlation  $r_L$ ).

**Table 1**  
**Summary of exon, transcript and gene level performance.**

	Gene																		
	Exon						Transcript												
	<i>H. sapiens</i>		<i>D. melanogaster</i>		<i>C. elegans</i>		<i>H. sapiens</i>		<i>D. melanogaster</i>		<i>C. elegans</i>		<i>H. sapiens</i>		<i>D. melanogaster</i>		<i>C. elegans</i>		
Sensitivity	Precision	Sensitivity	Precision	Sensitivity	Precision	Sensitivity	Precision	Sensitivity	Precision	Sensitivity	Precision	Sensitivity	Precision	Sensitivity	Precision	Sensitivity	Precision	Sensitivity	Precision
AUGUSTUS all	81.0%	77.47%	82.01%	86.03%	91.76%	76.26%	39.26%	42.66%	61.14%	59.28%	44.09%	56.16%	42.71%	73.03%	65.92%	72.35%	46.62%		
AUGUSTUS high	81.0%	84.15%	n.a.	n.a.	91.48%	83.55%	16.27%	16.27%	n.a.	59.21%	51.96%	56.16%	49.78%	n.a.	n.a.	72.26%	53.33%		
AUGUSTUS no RNA	65.55%	50.73%	71.12%	77.83%	86.51%	68.56%	3.52%	23.30%	42.22%	42.25%	31.53%	13.38%	8.03%	41.21%	42.22%	52.77%	31.53%		
Cufflinks	73.45%	79.57%	71.73%	83.70%	78.38%	84.84%	16.03%	36.18%	46.32%	39.75%	37.53%	53.55%	20.36%	56.37%	50.06%	48.04%	42.88%		
Exonerate all	74.59%	88.34%	73.32%	86.35%	76.99%	85.55%	17.22%	38.51%	51.44%	43.82%	40.02%	55.53%	28.60%	58.83%	51.44%	52.90%	40.02%		
Exonerate high	72.72%	92.46%	69.54%	92.50%	75.75%	89.27%	13.23%	30.98%	66.02%	39.98%	50.42%	50.28%	41.92%	54.54%	66.02%	49.83%	50.42%		
GSTRUCT	70.63%	96.42%	59.66%	95.60%	71.48%	91.85%	14.10%	28.21%	76.77%	39.03%	59.92%	54.02%	59.85%	49.10%	76.77%	48.72%	59.92%		
mGene	78.40%	50.96%	75.49%	91.06%	90.99%	78.64%	8.98%	40.00%	67.02%	55.27%	50.73%	41.35%	9.00%	67.17%	67.05%	69.01%	50.73%		
mGene graph	77.27%	73.14%	73.03%	90.92%	83.26%	81.96%	14.90%	40.00%	64.82%	55.78%	54.68%	48.17%	21.42%	67.78%	70.56%	68.88%	56.18%		
mTim	64.86%	90.42%	62.35%	90.40%	77.16%	88.48%	10.14%	20.59%	40.70%	31.86%	37.72%	35.78%	36.95%	34.09%	53.27%	38.96%	46.54%		
NextGeneid	67.36%	89.58%	70.53%	89.19%	78.45%	86.60%	11.30%	36.30%	62.23%	37.04%	44.55%	43.27%	36.99%	64.27%	62.23%	46.19%	44.55%		
NextGeneidAS	66.65%	87.32%	73.97%	83.76%	77.91%	82.80%	11.96%	40.05%	49.64%	38.45%	35.17%	42.73%	27.73%	66.30%	49.64%	46.71%	35.17%		
NextGeneidAS ab-initio	66.68%	85.56%	73.95%	83.66%	77.85%	82.76%	11.94%	40.05%	49.61%	38.45%	35.41%	42.63%	26.60%	66.25%	49.61%	46.71%	35.41%		
Oases	62.02%	40.55%	64.31%	51.05%	66.77%	61.04%	10.04%	33.99%	18.61%	32.98%	17.39%	34.05%	4.74%	55.47%	25.39%	39.70%	22.24%		
Transomics all	76.84%	54.35%	73.74%	74.03%	93.90%	69.92%	8.41%	35.11%	44.78%	58.10%	39.32%	32.16%	13.28%	62.71%	44.78%	72.52%	39.32%		
Transomics high	76.14%	74.64%	73.63%	79.67%	93.43%	79.73%	8.41%	35.11%	50.38%	58.08%	48.32%	32.16%	21.74%	62.69%	50.38%	72.50%	48.32%		
Trembly all	75.45%	83.44%	n.a.	n.a.	n.a.	n.a.	21.82%	n.a.	n.a.	n.a.	n.a.	60.53%	38.37%	n.a.	n.a.	n.a.	n.a.		
Trembly high	64.30%	90.24%	n.a.	n.a.	n.a.	n.a.	12.84%	n.a.	n.a.	n.a.	n.a.	43.60%	41.80%	n.a.	n.a.	n.a.	n.a.		
Tronter	49.02%	37.22%	29.15%	27.41%	39.43%	32.92%	4.27%	9.37%	4.79%	7.24%	1.96%	13.92%	6.48%	15.69%	10.47%	8.97%	4.96%		
Velvet	62.05%	59.56%	58.59%	58.75%	67.44%	72.86%	6.81%	24.09%	22.63%	31.77%	27.15%	21.35%	7.26%	42.39%	24.81%	39.54%	28.09%		
Velvet + AUGUSTUS	38.97%	90.68%	n.a.	n.a.	61.49%	69.46%	7.17%	n.a.	n.a.	0.00%	0.00%	27.44%	45.35%	n.a.	n.a.	0.00%	0.00%		
iReckon	73.24%	83.26%	82.93%	80.44%	87.39%	83.71%	27.78%	57.05%	51.13%	68.45%	44.85%	71.61%	35.77%	83.09%	51.13%	79.25%	44.85%		
SLIDE all	72.58%	84.10%	89.08%	95.56%	88.52%	97.34%	8.78%	52.80%	31.26%	42.85%	13.34%	29.08%	34.60%	82.02%	74.88%	51.92%	45.44%		
SLIDE high	54.75%	88.22%	82.31%	96.83%	82.97%	98.04%	8.65%	53.22%	58.60%	58.13%	51.78%	31.09%	31.62%	88.50%	80.58%	71.61%	62.66%		