

## Large-scale benchmarking of circRNA detection tools reveals large differences in sensitivity but not in precision

Peer-reviewed author version

Vromman, Marieke; Anckaert, Jasper; Bortoluzzi, Stefania; Buratin, Alessia; Chen, Chia-Ying; Chu, Qinjie; Chuang, Trees-Juen; Dehghannasiri, Roozbeh; Dieterich, Christoph; Dong, Xin; Flicek, Paul; Gaffo, Enrico; Gu, Wanjun; He, Chunjiang; Hoffmann, Steve; Izuogu, Osagie; Jackson, Michael S.; Jakobi, Tobias; Lai, Eric C.; Nuytens, Justine; Salzman, Julia; Santibanez-Koref, Mauro; THAS, Olivier; Stadler, Peter; Eynde, Eveline Vanden; Verniers, Kimberly; Wen, Guoxia; Westholm, Jakob; Yang, Li; Ye, Chu-Yu; Yigit, Nurten; Yuan, Guo-Hua; Zhang, Jinyang; Zhao, Fangqing; Vandesompele, Jo & VOLDERS, Pieter-Jan (2023) Large-scale benchmarking of circRNA detection tools reveals large differences in sensitivity but not in precision. In: NATURE METHODS, 20 (8), p. 1159 -1169.

DOI: 10.1038/s41592-023-01944-6

Handle: <http://hdl.handle.net/1942/40774>

# Large-scale benchmarking of circRNA detection tools reveals large differences in sensitivity but not in precision

Marieke Vromman<sup>1</sup>, Jasper Anckaert<sup>1</sup>, Stefania Bortoluzzi<sup>2</sup>, Alessia Buratin<sup>2</sup>, Chia-Ying Chen<sup>3</sup>, Qinjie Chu<sup>4</sup>, Trees-Juen Chuang<sup>3</sup>, Roozbeh Dehghannasiri<sup>5</sup>, Christoph Dieterich<sup>6</sup>, Xin Dong<sup>7</sup>, Paul Flicek<sup>8</sup>, Enrico Gaffo<sup>2</sup>, Wanjun Gu<sup>9</sup>, Chunjiang He<sup>7</sup>, Steve Hoffmann<sup>10</sup>, Osagie Izuogu<sup>8</sup>, Michael S. Jackson<sup>11</sup>, Tobias Jakobi<sup>12</sup>, Eric C. Lai<sup>13</sup>, Justine Nuytens<sup>1</sup>, Julia Salzman<sup>5</sup>, Mauro Santibanez-Koref<sup>11</sup>, Peter Stadler<sup>14</sup>, Olivier Thas<sup>15</sup>, Eveline Vanden Eynde<sup>1</sup>, Kimberly Verniers<sup>1</sup>, Guoxia Wen<sup>16</sup>, Jakub Westholm<sup>17</sup>, Li Yang<sup>18</sup>, Chu-Yu Ye<sup>4</sup>, Nurten Yigit<sup>1</sup>, Guo-Hua Yuan<sup>19</sup>, Jinyang Zhang<sup>20</sup>, Fangqing Zhao<sup>20</sup>, Jo Vandesompele<sup>1\*</sup>, Pieter-Jan Volders<sup>1\*</sup>

\* These authors jointly supervised this work.

<sup>1</sup> OncoRNALab, Cancer Research Institute Ghent (CRIG), Department of Biomolecular Medicine, Ghent University, Belgium

<sup>2</sup> Department of Molecular Medicine, University of Padova, Italy

<sup>3</sup> Genomics Research Center, Academia Sinica, Taiwan

<sup>4</sup> Institute of Crop Science & Institute of Bioinformatics, Zhejiang University, China

<sup>5</sup> Department of Biomedical Data Science and of Biochemistry, Stanford University, USA

<sup>6</sup> Klaus Tschira Institute for Integrative Computational Cardiology, Department of Internal Medicine III, University Hospital Heidelberg, German Center for Cardiovascular Research (DZHK), Germany

<sup>7</sup> School of Basic Medical Science, Department of Medical Genetics, Wuhan University, China

<sup>8</sup> EMBL-EBI, UK

<sup>9</sup> Collaborative Innovation Center of Jiangsu Province of Cancer Prevention and Treatment of Chinese Medicine, School of Artificial Intelligence and Information Technology, Nanjing University of Chinese Medicine, China

<sup>10</sup> Computational Biology Group, Leibniz Institute on Aging - Fritz Lipmann Institute (FLI), Jena, Germany

<sup>11</sup> Biosciences Institute, Faculty of Medical Sciences, Newcastle University, UK

<sup>12</sup> Translational Cardiovascular Research Center, University of Arizona - College of Medicine Phoenix, USA

<sup>13</sup> Developmental Biology Program, Sloan Kettering Institute, New York, USA

<sup>14</sup> Bioinformatics Group, Department of Computer Science, and Interdisciplinary Center for Bioinformatics, Universität Leipzig, Germany

<sup>15</sup> Data Science Institute, I-Biostat, Hasselt University, Belgium

<sup>16</sup> State Key Laboratory of Bioelectronics, School of Biological Sciences and Medical Engineering, Southeast University, China

<sup>17</sup> Dept of Biochemistry and Biophysics, National Bioinformatics Infrastructure Sweden, Science for Life Laboratory, Stockholm University, Sweden

<sup>18</sup> Center for Molecular Medicine, Children's Hospital, Fudan University and Shanghai Key Laboratory of Medical Epigenetics, International Laboratory of Medical Epigenetics and Metabolism, Ministry of Science and Technology, Institutes of Biomedical Sciences, Fudan University, China

<sup>19</sup> CAS Key Laboratory of Computational Biology, Shanghai Institute of Nutrition and Health, University of Chinese Academy of Sciences, Chinese Academy of Sciences, China

<sup>20</sup> Beijing Institutes of Life Science, Chinese Academy of Sciences, China

## ABSTRACT

The detection of circular RNA molecules (circRNAs) is typically based on short-read RNA sequencing data processed by computational tools. During the last decade, a plethora of such tools has been developed, but a systematic comparison with orthogonal validation is missing. Here, we set up a circRNA detection tool benchmarking study, in which 16 tools detected over 315,000 unique circRNAs in three deeply sequenced human cell types. Next, 1,516 predicted circRNAs were validated using three orthogonal methods. Generally, tool-specific precisions are high and similar (median of 98.8%, 96.3%, and 95.5% for qPCR, RNase R, and amplicon sequencing, respectively) whereas the sensitivity and number of predicted circRNAs (ranging from 1,372 to 58,032) are the most significant tool differentiators. **Of note, precision values are lower when evaluating low-abundance circRNAs.** We also demonstrate the complementarity of tools through the increase in detection sensitivity by considering the union of highly-precise tools while keeping the number of false discoveries low. Finally, recommendations are put forward for circRNA detection and validation.

## MAIN TEXT

### Introduction

Circular RNAs (circRNAs) are a class of non-coding RNA molecules numerous present in humans and other eukaryotic species. For a long time, circRNAs were regarded as unimportant byproducts of splicing. However, since the advancement of RNA sequencing technologies and the development of circRNA detection bioinformatics pipelines, there has been a significant increase in circRNA research, with a compound annual growth rate of scientific publications of 58% over the last five years (Figure 1A) (1).

Although an *in vivo* function for most circRNAs remains unknown and functional analyses are typically restricted to *in vitro* experiments, some circRNAs have been linked to specific diseases, including cancer. CircRNAs have also been reported to be more stable than linear transcripts due to the absence of a free 5' or 3' end to be recognized by exonucleases (1). In line with this, a higher fraction of circRNA relative to linear RNA has been observed in a wide range of human biofluids, which makes them interesting biomarker candidates, with the potential to be used for minimally-invasive tests for diagnosis or response monitoring (2). Wang *et al.* reviewed 112 differentially expressed circRNAs in various biofluids from patients with different cancer types (3). Furthermore, 15 clinical trials incorporating circRNAs as disease biomarkers have been initiated (ClinicalTrials.gov, accessed on 20/10/2022).

Eukaryotic circRNAs are formed through a process called back-splicing, where the 5' end of an RNA molecule forms a covalent bond with its own 3' end, forming a circular molecule with a characteristic back-spliced junction (BSJ) sequence (Figure 1B) (1). CircRNAs comprise of one or multiple exons, and analogous to linear RNA, there is alternative splicing of circRNAs, where circRNAs with the same BSJ sequence may have a different exon (and/or intron) composition (1).

In a targeted manner, circRNAs can be quantified with RT-qPCR (reverse transcription quantitative polymerase chain reaction) using BSJ-spanning primer pairs to amplify the region flanking the BSJ (Figure 1C). These primer pairs are divergent (facing away from each other) when hybridizing to the linear host transcripts and can therefore only amplify the circRNA (4). However, false positives resulting from alignment ambiguity, repeat sequences, trans-splicing, or RT template-switching artifacts have been described (5, 6). In all these cases, a linear RNA molecule is formed with the same exon orientation and, therefore, the same sequence as the circRNA BSJ. To prevent false positive circRNA identifications, linear RNA is often digested with the exonuclease ribonuclease R (RNase R) followed by RT-qPCR. RNase R typically degrades linear RNA, whereas circRNAs are generally not affected. Of note, it has been suggested that long circRNAs may be somewhat sensitive to RNase R degradation and various challenges of validating circRNAs have been recognized (7, 8).

In general, high-throughput or exploratory circRNA detection is performed using bioinformatics approaches that analyze total RNA sequencing data. For this, the RNA sequencing reads are first mapped against a reference genome. The unmapped reads are subsequently used to identify BSJ-spanning reads that map divergently (in reverse order) on the linear genome (Figure 1D).

Over the last decade, numerous computational circRNA detection tools have been developed and tested. Whereas multiple sets of circRNA detection tools using a bioinformatics approach have been compared (often when a novel tool is published) (9–16), a systematic and comprehensive evaluation of many circRNA detection tools using an orthogonal validation method is still missing. In our benchmarking study, we aimed to evaluate all currently available circRNA detection tools with an orthogonal approach using RT-qPCR, RNase R, and amplicon sequencing (Figure 2A). Our study highlights that although the precision of the tools is generally excellent, their sensitivities are highly variable.

## Results

### CircRNA detection tools predict a wide variety of circRNAs

#### *CircRNA detection tools differ in detection strategies and filtering*

For this study, 16 different circRNA detection tools were included: CIRCexplorer3 (17), CirComPara2 (11), circRNA\_finder (18), circseq\_cup (19), CircSplice (20), circTools (21), CIRI2 (22), CIRIquant (23), ecircscreen (unpublished tool), find\_circ (24), KNIFE (15), NCLscan (25), NCLcomparator (26), PFv2 (27), Sailfish-cir (28), and segemehl (29) (Table 1, Supplementary Table 1). CircRNA detection tools differ in their circRNA detection approach (including strand assignment), reliance on linear annotation, and filtering methods. CircRNAs can be detected from RNA sequencing data using the pseudo-reference-based approach (also called the candidate-based approach) or the fragmented-based approach (also called the segmented-read-based approach) (12, 14). The former approach uses a reference list of potential BSJ sequences, often based on all possible combinations of known annotated exons within a gene. This approach is therefore limited to species with annotated genomes and to previously annotated genes and will only detect circRNAs that use the same splicing sites as the linear RNAs. The latter approach splits unmapped sequencing reads into shorter sequences and remaps these against the reference genome. Lastly, integrative tools, such as CirComPara2 and ecircscreen, combine the results of multiple tools.

#### *The number of detected circRNAs differs greatly among tools*

A total of 315,312 unique circRNA predictions (corresponding to 1,137,099 unique circRNA/strand/tool/sample tuples) were detected using 16 different tools based on deeply sequenced total RNA from three human cancer cell lines (Supplementary Table 2, because of large file size, only available on [https://github.com/OncoRNALab/circRNA\\_benchmarking](https://github.com/OncoRNALab/circRNA_benchmarking)). The circRNA detection tools were run by their developers (details in Methods and Supplementary Notes). There is a striking almost 40-fold difference between the tool with the highest number of predicted circRNAs (circseq\_cup with 58,032 circRNAs) and the tool with the lowest number of predicted circRNAs (segemehl with 1,372 circRNAs) for one of the cell lines (Figure 2B shows results for HLF cells, similar results for the other cell lines are shown in Supplementary Figure 1).

#### *Most circRNAs are characterized by low BSJ counts*

CircRNA abundance is reflected by the BSJ count, which is the number of reads uniquely assigned to a given circRNA. The majority of circRNAs (86.6%) are detected with a BSJ count below 5 (Figure 2B), with only 46.1% of the detected circRNAs being observed with at least 2 BSJ counts (detailed distribution in Supplementary Figure 2). To increase confidence, circRNA\_finder and segemehl filtered their results to report only circRNAs with a BSJ count of at least 5, and CirComPara2 and KNIFE filtered for circRNAs with a BSJ count of at least 2. CircTools filtered circRNAs with at least 2 counts in at least 2 samples. Of note, Sailfish-cir does not report raw BSJ counts, but transcripts per million (TPM) instead. The similarity of circRNA BSJ counts between tool pairs is reasonable, according to regression analysis (linear models with median  $r^2 = 0.86$ , median slope = 0.70, all p-values < 0.001, Supplementary Figure 3).

#### *CircRNA detection tools predict different sets of circRNAs*

Half of all circRNAs in this study (49.9%) are reported only by one tool, which is largely due to circseq\_cup's high number of uniquely predicted circRNAs (Figure 2C for HLF cells, similar results for the other cell lines are shown in Supplementary Figure 4). The overlap of circRNA predictions among different tools is visualized in a heat map for each cell line in Supplementary Figure 5. Out of 16 circRNA detection tools, 8 exclusively report circRNAs flanked by canonical splice sites (with an AGNGT pattern, where AG is the splice acceptor, N represents the circRNA sequence in between, and GT is the splice donor) (Figure 2D). CirComPara2, circseq\_cup, Sailfish-cir, and segemehl do not report circRNA strand orientation, explaining

most of the ACNCT patterns (reverse complement of AGNGT), as all their predicted circRNAs were automatically assigned to the positive strand to retrieve the surrounding splicing sequence. Two-thirds of all predicted circRNAs in this study (68.5%) are novel compared to a set of previously reported circRNAs extracted from 13 published circRNA databases (Circ2Disease, circad, CircAtlas, circbank, circBase, CIRCpediav2, CircR2disease, CircRiC, circRNADb, CSCD, exoRBase, MiOncoCirc, and TSCD) (Supplementary Figure 6) (30). Of note, approximately half of these novel circRNA candidates originate solely from circseq\_cup. Looking at the tools individually, circseq\_cup, KNIFE, NCLscan, and NCLcomparator report a higher number of novel circRNAs (87.8%, 53.9%, 53.4%, 53.3%, respectively) compared to the other tools (median 19.7%, interquartile range (IQR) 4.9-34.8%). Tools were further compared based on the predicted circRNA length, strand information, correspondence to linear annotation, and predicted exon composition (Supplementary Data 1-4, Supplementary Figure 7-10). No notable differences were observed among tools, except for CIRI2 and PFV2 having a higher number of circRNAs for which no canonical linear annotation match was found compared to the other circRNA detection tools (*i.e.*, the BSJ position of the circRNA did not match any known intron-exon splicing position based on the canonical transcripts from Ensembl GRCh38.103). Across all tools, 53.7% of circRNAs uniquely match one canonical linear transcript, 10.3% match more than one canonical transcript, and 35.9% do not match any canonical transcript. CircRNAs were found for 17,461 different canonical human transcripts, demonstrating the pervasive nature of back-splicing (28.9% of canonical transcripts from Ensembl GRCh38.103). Of note, this is an underestimation, as for 46.3% of circRNAs, no (unique) annotation match could be found.

#### CircRNA validation with empirical methods

##### *CircRNA primer design inherently introduces a selection bias*

Based on previous experiments (Supplementary Data 5, Supplementary Figure 11), for each tool, we aimed to select 80 random *high-abundance* circRNAs with a BSJ count of at least 5 and 20 random *low-abundance* circRNAs with a BSJ count below 5. Importantly, the precision values for both abundance groups (described in the following paragraphs) cannot be directly compared due to the differences in the sample size. Of note, circRNA primer design inherently introduces a bias caused by the discarding of primer pairs (and therefore circRNAs) with predicted off-target amplification (Supplementary Data 6, Supplementary Figure 12). A selection of 1,560 circRNAs was obtained (Supplementary Table 3, BSJ count distribution in Supplementary Figure 13). As some circRNAs were selected more than once (by chance, for different tools or in different cell lines), the total number of unique circRNAs/sample pairs was 1,516, from here on termed 'selected circRNAs' (detailed description in Supplementary Figure 14). Furthermore, a second bias is introduced within the group of *low-abundant* circRNAs, as three tools (CirComPara2, circTools, and KNIFE) filter circRNAs with a BSJ count of at least two, whereas all other tools in this category also allow circRNAs with a BSJ count of one.

##### *High BSJ detection precision using RT-qPCR validation*

Of the 1,516 selected circRNAs, 1,479 (97.6%) could be validated with RT-qPCR, *i.e.*, the primer pair flanking the BSJ site resulted in a detectable amplicon. For the *low-abundance* circRNAs there is some variation in the tool-specific precision values (median 95.0%, range 80.0-100%), which is expected. *High-abundance* circRNAs have high RT-qPCR precision values for most tools (median 98.8%, range 90.0-100%) (Figure 3A; the cumulative plot of the RT-qPCR precision in function of the BSJ count is shown in Supplementary Figure 15). It is important to note that RT-qPCR-based validation is the net result of a successful primer pair and the actual presence of a sufficiently abundant circRNA in the amount of RNA tested.

### ***RNase R treatment degrades one in sixteen predicted circRNAs***

RNase R was used as a second, more stringent validation approach. RNase R selectively degrades linear transcripts, ensuring the RT-qPCR primers amplify a circular molecule. For 112 out of 1,516 selected circRNAs (7.4%), RNase R treatment could not be evaluated, as their abundance in the untreated sample was too low, leaving no room to confirm RNase R degradation in the event of a false positive circRNA (hence labeled as NAs). In the remaining set of 1404 predicted circRNAs, 1319 circRNAs (93.9%) could be successfully validated using RT-qPCR on RNase R treated RNA. For most tools, high RNase R precision values were observed for *high-abundance* circRNAs (median 96.3%, range 74.0-100%). PFv2 displays the lowest precision (74.0%). For *low-abundance* circRNAs, lower precision values were observed (median 86.7%, range 50.0-100%) (Figure 3A; the cumulative plot of the RNase R precision in function of the BSJ count is shown in Supplementary Figure 15). Of note, the number of circRNAs per tool in this bin is lower than the original 20 that were selected, as more circRNAs were excluded due to too low abundance (resulting in only 10-18 circRNAs per tool, with a median of 14 circRNAs). A comparison with matched RNase R treated and untreated sequencing data is described in Supplementary Data 7 and shows that the RNase R precision calculated from sequencing results is mostly high and similar among tools, with PFv2 having the lowest precision value (Supplementary Figures 16-19, Supplementary Tables 4 and 5, because of large file size, only available on [https://github.com/OncoRNALab/circRNA\\_benchmarking](https://github.com/OncoRNALab/circRNA_benchmarking)).

### ***Amplicon sequencing is the most stringent validation method***

The RT-qPCR amplicons of the untreated RNA were sequenced for further validation of the circRNAs. A random subset of circRNAs (1,179/1,516, 77.8%) was included in the amplicon sequencing experiment, resulting in a variable number of circRNAs per tool instead of 20 or 80 for *low-abundance* and *high-abundance* circRNAs, respectively (range 11-20 or 54-74 circRNAs per tool, respectively). For the remaining 1,179 circRNAs, 1,014 circRNAs (86.0%) could be readily validated with amplicon sequencing, i.e., the majority of reads aligned to the expected BSJ sequence. Most tools have similar amplicon sequencing precision values for *high-abundance* circRNAs (median 95.5%, range 30.0-100%), with PFv2 displaying a very low (30.0%) amplicon sequencing precision value. Of note, as PFv2 was developed to retain repeat sequences, it is expected to result in more false positives. The most obvious are caused by linear read-through between exons in neighboring tandemly repeated gene clusters/interspersed repeats, and these tend to be abundant. For *low-abundance* circRNAs, performance is more diverse, with generally lower amplicon sequencing precision (median 73.3%, range 17.6-94.1%) (Figure 3A; the cumulative plot of the amplicon sequencing on-target amplification rate and the cumulative plot of the amplicon sequencing precision in function of the BSJ count are shown in Supplementary Figures 20 and 15, respectively).

### ***Different validation methods should be used in concert***

Although the three validation strategies were used independently, it is interesting to evaluate to what extent they support each other (Figure 3B, Supplementary Figures 21 and 22). Considering 1,103 circRNAs for which all three validation results are available, 957 circRNAs (86.8%) pass all validation methods, 128 circRNAs (11.6%) fail one or two of the validation methods, and 18 circRNAs (1.6%) fail all three validation methods. These observations show that orthogonal validation with different empirical approaches is important to compensate for their inherent limitations. It is beyond the scope of this study to investigate why there are some discrepancies among the validation results (some hypotheses are considered in the Discussion). First, they are rare (for most circRNAs, the different methods completely agree). Second, the same methods are used to compare the tools, whereby no tool should be favored over the other.

The three orthogonal validation methods were combined to label each circRNA as a true or false positive result and the compound precision was calculated for each tool. Similar to the separate precision values, the compound precision is high and similar for most tools when looking at *high-abundance* circRNAs (median 93.1%, range 27.1-98.3%, IQR 90.5-95.3%),

and lower and more variable for *low-abundance* circRNAs (median 63.6%, range 5.9-88.2%, IQR 53.8-76.5%) (Supplementary Figures 15 and 23).

#### *CircRNA detection tools differ greatly in sensitivity*

Tool sensitivity was evaluated using two different methods. First, sensitivity was calculated based on the total number of true positive circRNAs ( $n = 957$ ) (Supplementary Figures 24 and 25). Of note, this sensitivity metric should be used with caution as it is based on a biased set of circRNAs collected from the 100 circRNAs selected per tool, which overlap and are not a representative random sample of all circRNAs (see Methods). Second, the theoretical number of true positive circRNAs for each tool was computed by multiplying the total number of detected circRNAs and the compound precision (*i.e.*, the extrapolated sensitivity) (Figure 3C for HLF cells, similar results for the other cell lines are shown in Supplementary Figure 26). There is a significant positive correlation between both sensitivity values (Spearman rank correlation of 0.84 with  $p$ -value  $< 0.001$ ,  $S = 58$ , for *low-abundance* circRNAs, and a Spearman rank correlation of 0.80 with  $p$ -value  $< 0.001$ ,  $S = 113$ , for *high-abundance* circRNAs). Both methods show great variability in tool sensitivity, with a median sensitivity of 75.1% (range 29.7-87.1%) for *low-abundance* circRNAs and 65.7% (range 18.7-87.4%) for *high-abundance* circRNAs. To visualize the relationship between sensitivity and compound precision, a precision-recall (sensitivity) dot plot for all tools is shown in Supplementary Figure 27. All metrics described above ((compound) precisions and sensitivity) and the tool ranking for each metric are available in Supplementary Table 6. The user can easily filter and order the circRNA detection tools based on their preferences. Reproducibility evaluations were performed and are described in Supplementary Data 8-10 (Supplementary Figures 28-32).

#### **Evaluation of precision in function of circRNA annotation**

To compare precision values in function of circRNA annotation, we restrict the analyses to *high-abundance* circRNAs with information for all validation techniques. Furthermore, a strict validation definition was used, where all circRNAs failing for at least one technique were classified as unvalidated. CircRNAs previously described in databases have higher chances of getting validated (Chi-squared = 181.0, degrees of freedom (df) = 1,  $p$ -value  $< 0.001$ , odds ratio (OR) = 13.1). Nevertheless, false positive circRNAs according to our data are still present in multiple published databases (Supplementary Figures 33-35). For example, false positive circRNA chr6:47526627-47554766 (hg38, 0-based) is present in CircAtlas (as hsa-CD2AP\_0048) and in exoRBase (as exo\_circ\_65199). A difference in validation rate in function of the splicing pattern was observed, with better validation of circRNAs surrounded by canonical splice sites (Chi-squared = 45.4, df = 1,  $p$ -value  $< 0.001$ , OR = 5.0). Similarly, circRNAs that originate from a region with an annotated linear transcript have higher validation rates (Chi-squared = 185.8, df = 1,  $p$ -value  $< 0.001$ , OR = 17.1). Surprisingly, single-exon circRNAs displayed significantly lower validation rates than multi-exon circRNAs (Chi-squared = 20.0, df = 1,  $p$ -value  $< 0.001$ , OR = 3.8). Lastly, while tools with a 'candidate-based' approach seem more precise than tools using the 'segmented read-based' approach (Chi-squared = 9.4, df = 1,  $p$ -value = 0.0022, OR = 2.6), we cannot be sure that these results are not confounded by other algorithmic differences.

#### **Evaluation of sensitivity in function of circRNA annotation**

There is a significantly higher sensitivity for tools reporting circRNAs surrounded by canonical splice sites, resulting in a median difference in sensitivity of 38.5% (two-sided Mann-Whitney  $U = 55$ ,  $p$ -value = 0.0022, large effect size of 0.78, 95% CI [0.56 - 0.85],  $n_1 = 11$ ,  $n_2 = 5$ , only *high-abundance* circRNAs). However, no link could be found between sensitivity and tool approach, use of linear annotation, strand annotation method, or BSJ count filtering.

### Evaluation of tool combinations to improve performance

For the combination of two or more tools, both the intersection and the union have been proposed (11) (Supplementary Tables 7 and 8). While not evaluated here, the increased time and resource consumption should also be taken into account when considering the use of multiple tools. A list of the top-performing combinations is available in Supplementary Table 9 and can be used as a reference.

#### ***A circRNA predicted by two tools can be a false result***

Figure 4A shows that circRNAs uniquely detected by a single tool generally have lower precision values. In line with this, circRNAs detected by at least two tools have a higher chance of getting validated (Chi-squared = 333.1, df = 1, p-value < 0.001, OR = 53.8). On the other hand, out of 1,380 unique circRNAs detected by at least two tools, 7 circRNAs (0.5%) failed all three validation methods, and 137 (9.9%) failed at least one of the validation methods (Supplementary Figure 36), illustrating that the practice of using the intersection is not a guarantee to avoid false positive results.

#### ***The union of tools increases the number of true circRNAs***

To maximize detection sensitivity and maintain precision, we evaluated the union of pairs or triples of circRNA detection tools. Generating all possible combinations of the better tools with individual compound precision  $\geq 90\%$  for *high-abundance* circRNAs ( $n = 12$  tools) consistently results in higher detection sensitivity while maintaining a high weighted precision value. The median increase in the number of detected circRNAs is 37.0% (IQR 16.5-129.7%) and 79.6% (IQR 33.1-215.7%) for combinations of 2 or 3 tools, respectively. In other words, when combining very precise tools, the number of false positives does not counteract the gain in additional true positives. A subset of tool combinations with an increase of at least 1,000 circRNAs is shown in Figure 4B (shown for HLF cells, similar results for the other cell lines are shown in Supplementary Figure 37; the combo of three tools is shown in Supplementary Figure 38). One obvious consideration when selecting two different tools is their circRNA detection approach, their reliance on linear annotation, and their filtering methods. For example, when combining two tools with a different detection approach (pseudo-reference-based and fragmented-based approach), the median increase in the number of detected circRNAs is 61.1%, compared to 35.4% for two tools with the same detection approach (two-sided Mann-Whitney U = 58554.5, p-value < 0.001, small effect size of 0.16, 95% CI [0.08 - 0.23],  $n_1 = 336$ ,  $n_2 = 294$ ). Similarly, when combining two tools with the same splice site settings (both canonical or both non-canonical), the median increase in the number of detected circRNAs (32.6%) is significantly smaller compared to combining two tools with different splice site settings (one canonical and one non-canonical) (76.2%) (two-sided Mann-Whitney U = 65356.5, p-value < 0.001, small effect size of 0.28, 95% CI [0.21 - 0.35],  $n_1 = 300$ ,  $n_2 = 330$ ). A similar analysis for the combination of tools that rely or do not rely on linear annotation was not significant.

## **Discussion**

Multimodal orthogonal validation of bioinformatics tools that predict circular RNAs from total RNA sequencing data is currently lacking. Hence, their precision and sensitivities are unknown and scientific data is confounded with false positive and false negative predictions. To accommodate this lacune, we set up a large-scale international collaborative circRNA detection tool benchmarking study (Figure 2A). First, a deeply sequenced total RNA sequencing dataset was processed by the developers of 16 different circRNA detection tools. Next, 3 empirical validation strategies were used to evaluate a random selection of 1,560 circRNAs representing each tool: 1) RT-qPCR to determine if the candidate circRNA BSJ (back-spliced junction) sequence was detectable; 2) RNase R treatment to confirm that the detected RNA was most likely circular and not linear; and 3) amplicon sequencing to confirm the circRNA BSJ sequence. Of note, both circRNA RT-qPCR and RNase R validation protocols were extensively validated (4, 31).

The precision values are similarly high among tools (Figure 3A), especially when considering the subset of *high-abundance* circRNAs (with a BSJ count  $\geq 5$ ). In contrast, the number of predicted circRNAs and the sensitivity varies greatly among tools, in line with previous studies based on simulated data (11, 13, 14) (Supplementary Data 11, Supplementary Figures 39 and 40). The striking differences in sensitivity are in part dependent on the operator applying BSJ count filters.

The three validation methods each have their own strengths and biases, with conflicting results for several circRNAs (Figure 3B, discussed in detail in Supplementary Discussion 1). In total, 22 circRNAs are validated with qPCR and amplicon sequencing but are degraded by RNase R for at least 87.5% (*i.e.*, a difference of 3 cycles). A possible explanation could be that some *bona fide* circRNAs are susceptible to RNase R degradation (8) or that the primers amplify a mixture of circular and linear RNA. Another subset of 92 circRNAs pass RT-qPCR validation and RNase R validation but fail amplicon sequencing. These could be (repetitive) RNAs resistant to RNase R due to secondary structure, either internal or through base pairing with orthologs (8). These examples underscore the importance of using different validation methods to compensate for their intrinsic limitations and to increase the validation status confidence (as previously suggested in (8)).

Although long-read sequencing has been implemented to study full-length circRNAs (32–35), the bulk of currently available data remains short-read sequencing. Therefore, this benchmarking study evaluated circRNA detection tools for short-read sequencing data, which typically report circRNAs by their BSJ position (chr, start, end, strand). However, it remains unknown if the detected BSJ corresponds to one circRNA, or multiple alternatively spliced circRNAs with different exon/intron compositions. Henceforth, the prediction precision values reported here might be influenced by more than one circRNA with the same BSJ. As this study is focused on circRNA detection in short-read sequencing data, the internal circRNA composition was not evaluated. Furthermore, no distinction can be made between circRNAs on the positive strand or negative strand using RT-qPCR and amplicon sequencing (9.4% of circRNAs were reported to originate from different strands according to different tools).

Based on a pilot study (Supplementary Data 5, Supplementary Figure 11), a cut-off was set at BSJ count 5, as circRNAs under this cut-off approached the qPCR limit to reliably detect RNase R based degradation of falsely predicted circRNAs. While very deep sequencing of a large RNA input amount was performed, it is beyond the scope of this study to evaluate if the BSJ count should be reconsidered in function of sequencing depth. However, as the majority of predicted circRNAs have a BSJ count below 5, we decided to include at least a subset of these *low-abundance* circRNAs to calculate the corresponding prediction precision. It is no surprise that the precision values for *low-abundance* circRNAs are significantly lower compared to *high-abundance* circRNAs (Chi-squared = 76.7, df = 1, p-value < 0.001, OR = 3.8). This difference is likely due to the detection limits of the applied validation strategies in conjunction with the sampling bias of *low-abundance* analytes, and not due to inherently more false positive predictions for circRNAs with a lower count. Of note, it can be presumed that weakly expressed circRNAs are less relevant for both functional studies and biomarker research.

Focusing on *high-abundance* circRNAs, interesting associations were found between circRNA annotation and validation rates. As such, circRNAs had higher validation rates when they were detected by multiple tools, when they were previously reported in a circRNA database, when they were surrounded by canonical splice sites, and when they originate from a region with an annotated linear transcript. CircRNA detection tools with a 'candidate-based' approach are more precise than tools using the 'segmented read-based' approach, which is in line with the higher validation likelihood of circRNAs originating from known linear genes and surrounded by canonical splice sites.

Based on our study, we compiled a list of recommendations for circRNA detection and validation, and for the future development of circRNA detection tools and their performance evaluation (Table 2). Ideally, publicly available (spike-in) reference material (consisting of known synthetic circRNAs) should be used to benchmark existing and novel circRNA detection tools. However, such reference material is currently not available. As the main goal of this

study was to perform a neutral assessment of circRNA detection tool sensitivity and precision, the developers of the tools were asked to run the tools themselves. Therefore, execution time, memory usage, and ease of use could not be compared and were not assessed here. Furthermore, this study resulted in a circRNA resource containing > 315,000 circRNAs detected by different tools in three human cancer cell lines from different tissue origins and provides validation results for 1500 circRNAs that can be used as a reference for the development of new or improved circRNA detection tools. Finally, our study can also serve as an example framework for empirical validation of benchmarking results from other bioinformatics tools in the future.

## ACKNOWLEDGMENTS

We thank Steve Lefever for his contribution to primer design in the early stage of this project. The computational resources (Stevin Supercomputer Infrastructure) and services used in this work were provided by the VSC (Flemish Supercomputer Center), funded by Ghent University, FWO, and the Flemish Government – department EWI.

This work was supported by the Foundation Against Cancer grant STK F/2018/1,267 (M.V., J.V., P.J.V.), the Standup Against Cancer grant STIVLK2018000601 (J.V., P.J.V.), two Concerted Research Action of Ghent University grants BOF16/GOA/023 (J.V.), BOF/24J/2021/244 (M.V., J.V., P.J.V.), the Research Foundation - Flanders FWO grant 1253321N (P.J.V.), the Fondazione AIRC per la Ricerca sul Cancro Investigator Grant 2017 20052 (S.B.), the Italian Ministry of Education, Universities, and Research PRIN 2017 grant 2017PPS2X4\_003 (S.B.), the EU within the MUR PNRR “National Center for Gene Therapy and Drugs based on RNA Technology”, Project no. CN00000041 CN3 RNA (S.B.), the "HPC, big data and quantum computing" CN1 HPC (S.B.), the Department of Molecular Medicine of the University of Padova (S.B.), the National Health Research Institutes Taiwan grant NHRI-EX110-11011B1 (T.J.C.), the German Science Foundation grant DI 1501/13-1 (C.D.), the Wellcome Trust grant WT108749/Z/15/Z (P.F.), the Fondazione Umberto Veronesi Fellowship 2020 (E.G.), the German Federal Ministry of Education and Research grant BMBF 031L0106D (S.H.), the NIH grant R01-NS083833 (E.C.L.), the MSK Core Grant P30-CA008748 (E.C.L.), the German Federal Ministry of Education and Research grant BMBF 031L0164C (P.S.), the Knut and Alice Wallenberg Foundation as part of the National Bioinformatics Infrastructure Sweden at SciLifeLab (J.W.), the National Natural Science Foundation of China (NSFC) grant 31925011 (L.Y.), the Ministry of Science and Technology of China (MoST) grant 2021YFA1300503 (L.Y.), and the National Science Foundation of China (31871589) (C.Y.Y.).

## AUTHOR CONTRIBUTIONS

All authors have been ranked alphabetically in the author list according to their surname, except for the first author and the two shared last authors.

Validation co-author group: Marieke Vromman: methodology, software, validation, formal analysis, investigation, data curation, writing - original draft, visualization, funding acquisition; Jasper Anckaert: software; Justine Nuytens: investigation; Olivier Thas: methodology; Eveline Vanden Eynde: investigation; Kimberly Verniers: investigation; Nurten Yigit: investigation; Jo Vandesompele: conceptualization, methodology, writing - review & editing, supervision, project administration, funding acquisition; Pieter-Jan Volders: conceptualization, methodology, writing - review & editing, supervision, project administration, funding acquisition.

CircRNA prediction co-author group (contribution: formal analysis and writing - review & editing): Stefania Bortoluzzi, Alessia Buratin, Chia-Ying Chen, Qinjie Chu, Trees-Juen Chuang, Roozbeh Dehghannasiri, Christoph Dieterich, Xin Dong, Paul Flicek, Enrico Gaffo, Wanjun Gu, Chunjiang He, Steve Hoffmann, Osagie Izuogu, Michael S. Jackson, Tobias Jakobi, Eric C. Lai, Julia Salzman, Mauro Santibanez-Koref, Peter Stadler, Guoxia Wen, Jakob Westholm, Li Yang, Chu-Yu Ye, Guo-Hua Yuan, Jinyang Zhang, Fangqing Zhao

## **COMPETING INTERESTS**

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## **TABLES**

<b>tool</b>	<b>approach</b>	<b>circRNAs detected in ...</b>	<b>strand assignment*</b>	<b>splicing</b>	<b>BSJ count filter*</b>	<b>min circRNA length*</b>	<b>max circRNA length*</b>
CIRCexplorer3	segmented read-based	entire genome	based on linear annotation	AGNGT	none	none	none
CirComPara2	integrative <sup>x</sup>	entire genome	no strand reported	AGNGT, ACNCT	≥ 2	299	2,304,996
circRNA_finder	segmented read-based	entire genome	based on mapping to genome	AGNGT	≥ 5	200	100,000
circseq_cup	based on segemehl, with full-length circRNA assembly	entire genome	no strand reported	non-canonical	none	none	5,000
CircSplice	segmented read-based	known splice sites	based on linear annotation	AGNGT, ACNCT	none	78	none
circtools	segmented read-based	entire genome	based on mapping to genome	AGNGT	≥ 2 in ≥ 2 samples	31	1,000,000
CIRI2	segmented read-based	entire genome	based on GT-AG splice sites	AGNGT	none	135	200,000
CIRIquant	based on CIRI2, with improved quantification	entire genome	based on GT-AG splice sites	AGNGT	none	135	200,000
ecircscreen	integrative <sup>x</sup>	entire genome	based on consensus from tools	AGNGT	none	none	none
find_circ	segmented read-based	entire genome	based on mapping to genome	AGNGT	none	none	100,000
KNIFE	candidate-based	entire genome	based on linear annotation	non-canonical	≥ 2	none	1,000,000
NCLscan	candidate-based	known splice sites	based on linear annotation	non-canonical	none	100	none
NCLcomparator	filtered results of NCLscan	known splice sites	based on linear annotation	non-canonical	none	100	none
PFv2	segmented read-based	entire genome	based on mapping to genome	AGNGT, ACNCT	none	50	1,000,000
Sailfish-cir	based on CIRI2 v2.0.6	entire genome	no strand reported	AGNGT, ACNCT	no BSJ counts reported	135	200,000
segemehl	segmented read-based	entire genome	no strand reported	non-canonical	≥ 5	none	200,000

**Table 1** CircRNA detection tools with their circRNA detection approach, strand assignment approach, reliance on linear annotation, and filtering approach. <sup>x</sup> Integrative tools combine the results of multiple circRNA detection tools. This includes CirComPara2 (combining CIRCexplorer2 (v2.3.8), Segemehl (v0.3.4), CIRI2 (v2.0.6), DCC (v0.4.8), and find\_circ (v1.2), and the filtering all circRNAs detected by at least two methods) and ecircscreen (combining CIRI2 (v2.0.6), circRNA\_finder (v1.2), PFv2 (v2.0.0), find\_circ (v1.2), and CIRCexplorer (v1.1.10), and then filtering all circRNA detected by at least three methods). <sup>+</sup> Some tools did not report strand information for this study, but the (updated) circRNA tool might report circRNA strand information. <sup>\*</sup> The BSJ count, and minimum and maximum circRNA length filters, are the filters used for this specific study. The user can choose these parameters freely. Of note, the minimum and maximum length filters are based on the estimated circRNA length with introns, calculated by subtracting the start position from the end position of the BSJ. <sup>-</sup> Inferred based on publication and available code.

<b>circRNA detection</b>	<ol style="list-style-type: none"> <li>1. An orthogonal validation method must be used to validate a predicted circRNA; qPCR validation on its own is not sufficient, at least qPCR + RNase R treatment or preferably qPCR + amplicon sequencing should be used.</li> <li>2. Filtering based on a minimum BSJ count is recommended to increase the likelihood of successful empirical validation.</li> </ol>
<b>circRNA validation</b>	<ol style="list-style-type: none"> <li>3. For a precision-focused approach, the <i>intersection</i> of two tools with a high individual precision (for example <math>\geq 90\%</math>) should be used.</li> <li>4. For a sensitivity-focused approach, the <i>union</i> of two tools with a high individual precision (for example <math>\geq 90\%</math>) should be used.</li> <li>5. The choice of tools to be combined may be informed based on the tools' underlying principles (circRNA detection approach, reliance on linear annotation and canonical splicing, and filtering).</li> </ol>
<b>circRNA tool development</b>	<ol style="list-style-type: none"> <li>6. Tools should report the originating strand information, the BSJ count evidence, and the chromosomal start and end position of the BSJ.</li> </ol>
<b>circRNA tool validation</b>	<ol style="list-style-type: none"> <li>7. For evaluation of sensitivity, novel and updated tools are encouraged to use the empirically validated set of 957 true positive circRNAs.</li> <li>8. For evaluation of precision, a random set of 100 predicted circRNAs should be validated with empirical methods.</li> </ol>

**Table 2** CircRNA research recommendations

## FIGURE LEGENDS

**Figure 1** CircRNA scientific relevance, structure, and detection. **A.** Over the last decade, circRNA research has increased rapidly, as illustrated by the proportional growth of publications mentioning circRNA in Europe PubMed Central. **B.** CircRNAs are formed through back-splicing, which results in a circular molecule with a back-spliced junction (BSJ). Black boxes highlight the BSJ in the circRNA isoforms. **C.** CircRNAs can be detected with RT-qPCR using a BSJ-specific primer pair. The primer pair can only bind in a divergent manner (facing away from each other) to linear RNA, where no amplification will be possible, yet binds the circRNA in a convergent manner (facing towards each other), amplifying the BSJ sequence. **D.** Large-scale circRNA detection is typically performed using total RNA sequencing datasets and specialized computational tools. These tools identify BSJ-spanning reads, which map divergently (in reverse order) on the linear reference genome.

**Figure 2** CircRNA detection tools predict a wide variety of circRNAs **A.** This study consists of a circRNA detection phase and a circRNA validation phase. For the former, 16 circRNA

detection tools were used to predict circRNAs in three deeply sequenced cancer cell lines. For the latter, a set of circRNAs was selected per tool and validated using three orthogonal methods, generating tool-specific precision values for each method. This was also used to compute compound precision and both types of sensitivity values for each circRNA detection tool. **B.** The number of reported circRNAs differs greatly between tools (shown for HLF cells, similar results for the other cell lines are shown in Supplementary Figure 1). The tools are ordered according to the total number of predicted circRNAs. The vast majority of circRNAs are predicted with a BSJ count below 5 (in blue). Two tools, circRNA\_finder, and segemehl, filtered their results to report only circRNAs with a BSJ count of at least 5 (in orange). **C.** The majority of circRNAs (49.9%) are detected by only one tool. Circseq\_cup reports the largest set of unique circRNAs (shown for HLF cells, similar results for the other cell lines are shown in Supplementary Figure 4). A small set of 55 circRNAs is detected by all tools (column n\_db in Supplementary Table 2). **D.** CircRNA splice sites differ among circRNA detection tools. Most commonly, the canonical AGNGT pattern is observed, with AG being the splice acceptor, N the circRNA, and GT the splice donor. Circseq\_cup, CirComPara2, Sailfish-cir, and segemehl do not report strand information. To be able to retrieve a splicing sequence for the circRNAs from these tools, it was assumed the circRNA originated from the positive strand. This led to the ACNCT pattern (reverse complement of AGNGT), most probably from circRNAs that were assigned to the positive strand incorrectly. Lastly, there are some tools that also report a substantial number of circRNAs BSJ sequences with a GGNGG splicing pattern.

**Figure 3** The precision of circRNA detection tools is generally high and similar, whereas tools largely differ with respect to the number of predicted circRNAs. The plots are separated based on circRNA BSJ count below 5 (*low-abundance*, in blue, 20 circRNAs selected per tool) or a BSJ count of at least 5 (*high-abundance*, in orange, 80 circRNAs selected per tool). Sailfish-cir reports TPM (transcripts per million) values instead of BSJ counts, and is therefore depicted separately. As circRNA\_finder and segemehl do not report any circRNAs with a BSJ count < 5, these tools are not included in the blue bar plots. **A.** CircRNAs were validated using three different techniques: RT-qPCR detection, resistance to degradation by RNase R, and amplicon sequencing. *Low-abundance* circRNAs are in general more difficult to validate. Of note, the precision values for *low-abundance* circRNAs are based on a limited set of circRNAs. *High-abundance* circRNAs have good precision values for most tools and most validation methods. The error bars represent the 95% confidence intervals (CI). \* A set of circRNAs was excluded as they were too low abundant to assess their resistance to RNase R, resulting in a variable number of circRNAs per tool instead of 20 or 80 for *low-abundance* and *high-abundance* circRNAs, respectively (range 10-18 or 71-80 circRNAs per tool, respectively, details in Supplementary Table 6). A random subset of circRNAs was included in the amplicon sequencing experiment, resulting in a variable number of circRNAs per tool for amplicon sequencing validation as well (range 11-20 or 54-74 circRNAs per tool, respectively, details in Supplementary Table 6). **B.** The vast majority of circRNAs obtain the same verdict based on the three different validation methods. However, some circRNAs have conflicting results. For example, there are 13 circRNAs that are detectable by RT-qPCR but also are degraded upon RNase R treatment and for which the primers seem to amplify the wrong product. **C.** The compound precision value is used to compute the theoretical number of true positive circRNAs by multiplying it with the original number of circRNAs detected by that tool (*i.e.*, the extrapolated sensitivity) (shown for HLF, similar results for the other cell lines are shown in Supplementary Figure 26).

**Figure 4** The intersection or union of two circRNA detection tools decreases the number of false positives, or increases the overall number of detected circRNAs, respectively. **A.** CircRNAs detected by multiple tools generally have higher precision values. However, the often-used practice of using the intersection of two tools is not necessarily a guarantee for avoiding false positive results. **B.** By considering the union of two circRNA detection tools, the number of circRNAs can be significantly increased, whilst keeping the number of false positive predictions low (shown for the HLF cell line, similar results for the other two cell lines are

shown in Supplementary Figure 37). For the y-axis, the percentage of detected circRNAs is calculated by dividing the number of circRNA detected by that tool combination by the total number of predicted circRNAs for that sample taking the union of all tools (13,087 circRNAs for the HLF sample). For this analysis, the compound precision value of *high-abundance* circRNAs was used. Some circRNA detection tools are integrative and combine the results of multiple other tools. It is therefore assumed an integrative tool would have large similarities with its integrated tools. However, a difference in tool version and filtering can still give a different set of circRNAs. For example, CirComPara2 is an integrative tool that combines CIRCexplorer2, CIRI2, DCC, and find\_circ, nevertheless, the combination of CirComPara2 and CIRCexplorer3 still gives a significant increase in detected circRNAs (corresponding to 10% of all circRNA predictions for that cell line).

## REFERENCES

1. Kristensen,L.S., Andersen,M.S., Stagsted,L.V.W., Ebbesen,K.K., Hansen,T.B. and Kjems,J. (2019) The biogenesis, biology and characterization of circular RNAs. *Nat Rev Genet*, 20, 675–691.
2. Hulstaert,E., Morlion,A., Avila Cobos,F., Verniers,K., Nuytens,J., Vanden Eynde,E., Yigit,N., Anckaert,J., Geerts,A., Hindryckx,P., et al. (2020) Charting Extracellular Transcriptomes in The Human Biofluid RNA Atlas. *Cell Rep*, 33.
3. Wang,S., Zhang,K., Tan,S., Xin,J., Yuan,Q., Xu,H., Xu,X., Liang,Q., Christiani,D.C., Wang,M., et al. (2021) Circular RNAs in body fluids as cancer biomarkers: the new frontier of liquid biopsies. *Mol Cancer*, 20, 1–10.
4. Vromman,M., Yigit,N., Verniers,K., Lefever,S., Vandesompele,J. and Volders,P. (2021) Validation of Circular RNAs Using RT-qPCR After Effective Removal of Linear RNAs by Ribonuclease R. *Curr Protoc*, 1, e181.
5. Yu,C.Y., Liu,H.J., Hung,L.Y., Kuo,H.C. and Chuang,T.J. (2014) Is an observed non-co-linear RNA product spliced in trans, in cis or just in vitro? *Nucleic Acids Res*, 42, 9410.
6. Szabo,L. and Salzman,J. (2016) Detecting circular RNAs: Bioinformatic and experimental challenges. *Nat Rev Genet*, 17, 679–692.
7. Nielsen,A.F., Bindereif,A., Bozzoni,I., Hanan,M., Hansen,T.B., Irimia,M., Kadener,S., Kristensen,L.S., Legnini,I., Morlando,M., et al. (2022) Best practice standards for circular RNA research. *Nat Methods*, 19, 1208–1220.
8. Dodbele,S., Mutlu,N. and Wilusz,J.E. (2021) Best practices to ensure robust investigation of circular RNAs: pitfalls and tips. *EMBO Rep*, 22, 1–12.
9. Jakobi,T. and Dieterich,C. (2019) Computational approaches for circular RNA analysis. *Wiley Interdiscip Rev RNA*, 2019, e1528.
10. Hansen,T.B., Venø,M.T., Damgaard,C.K. and Kjems,J. (2015) Comparison of circular RNA prediction tools. *Nucleic Acids Res*, 44, e58.
11. Gaffo,E., Buratin,A., Dal Molin,A. and Bortoluzzi,S. (2022) Sensitive, reliable and robust circRNA detection from RNA-seq with CirComPara2. *Brief Bioinform*, 23, 1–12.
12. Jeck,W.R. and Sharpless,N.E. (2014) Detecting and characterizing circular RNAs. *Nat Biotechnol*, 32, 453–461.
13. Nguyen,D.T., Trac,Q.T., Nguyen,T.H., Nguyen,H.N., Ohad,N., Pawitan,Y. and Vu,T.N. (2021) Circall: fast and accurate methodology for discovery of circular RNAs from paired-end RNA-sequencing data. *BMC Bioinformatics*, 22.
14. Zeng,X., Lin,W., Guo,M. and Zou,Q. (2017) A comprehensive overview and evaluation of circular RNA detection tools. *PLoS Comput Biol*, 13, e1005420.

15. Szabo,L., Morey,R., Palpant,N.J., Wang,P.L., Afari,N., Jiang,C., Parast,M.M., Murry,C.E., Laurent,L.C. and Salzman,J. (2015) Statistically based splicing detection reveals neural enrichment and tissue-specific induction of circular RNA during human fetal development. *Genome Biol*, 16.
16. Song,X., Zhang,N., Han,P., Moon,B.S., Lai,R.K., Wang,K. and Lu,W. (2016) Circular RNA profile in gliomas revealed by identification tool UROBORUS. *Nucleic Acids Res*, 44, 1–12.
17. Ma,X.K., Wang,M.R., Liu,C.X., Dong,R., Carmichael,G.G., Chen,L.L. and Yang,L. (2020) CIRCexplorer3: A CLEAR Pipeline for Direct Comparison of Circular and Linear RNA Expression. *Genomics Proteomics Bioinformatics*, 17, 511–521.
18. Westholm,J.O., Miura,P., Olson,S., Shenker,S., Joseph,B., Sanfilippo,P., Celniker,S.E., Graveley,B.R. and Lai,E.C. (2014) Genome-wide Analysis of *Drosophila* Circular RNAs Reveals Their Structural and Sequence Properties and Age-Dependent Neural Accumulation. *Cell Rep*, 9, 1966–1980.
19. Ye,C.-Y., Zhang,X., Chu,Q., Liu,C., Yu,Y., Jiang,W., Zhu,Q.-H., Fan,L. and Guo,L. (2017) Full-length sequence assembly reveals circular RNAs with diverse non-GT/AG splicing signals in rice. *RNA Biol*, 14, 1055–1063.
20. Feng,J., Chen,K., Dong,X., Xu,X., Jin,Y., Zhang,X., Chen,W., Han,Y., Shao,L., Gao,Y., et al. (2019) Genome-wide identification of cancer-specific alternative splicing in circRNA. *Mol Cancer*, 18, 1–5.
21. Jakobi,T., Uvarovskii,A. and Dieterich,C. (2019) Circtools—a one-stop software solution for circular RNA research. *Bioinformatics*, 35, 2326–2328.
22. Gao,Y., Zhang,J. and Zhao,F. (2018) Circular RNA identification based on multiple seed matching. *Brief Bioinform*, 19, 803–810.
23. Zhang,J., Chen,S., Yang,J. and Zhao,F. (2020) Accurate quantification of circular RNAs identifies extensive circular isoform switching events. *Nat Commun*, 11, 90.
24. Memczak,S., Jens,M., Elefsinioti,A., Torti,F., Krueger,J., Rybak,A., Maier,L., Mackowiak,S.D., Gregersen,L.H., Munschauer,M., et al. (2013) Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature*, 495, 333–338.
25. Chuang,T.J., Wu,C.S., Chen,C.Y., Hung,L.Y., Chiang,T.W. and Yang,M.Y. (2016) NCLscan: Accurate identification of non-co-linear transcripts (fusion, trans-splicing and circular RNA) with a good balance between sensitivity and precision. *Nucleic Acids Res*, 44.
26. Chen,C.Y. and Chuang,T.J. (2019) NCLcomparator: Systematically post-screening non-co-linear transcripts (circular, trans-spliced, or fusion RNAs) identified from various detectors. *BMC Bioinformatics*, 20, 1–11.
27. Izuogu,O.G., Alhasan,A.A., Mellough,C., Collin,J., Gallon,R., Hyslop,J., Mastroianni,F.K., Ehrmann,I., Lako,M., Elliott,D.J., et al. (2018) Analysis of human ES cell differentiation establishes that the dominant isoforms of the lncRNAs RMST and FIRRE are circular. *BMC Genomics*, 19, 1–18.
28. Li,M., Xie,X., Zhou,J., Sheng,M., Yin,X., Ko,E.A., Zhou,T. and Gu,W. (2017) Quantifying circular RNA expression from RNA-seq data using model-based framework. In *Bioinformatics*. Vol. 33, pp. 2131–2139.
29. Hoffmann,S., Otto,C., Doose,G., Tanzer,A., Langenberger,D., Christ,S., Kunz,M., Holdt,L.M., Teupser,D., Hackermüller,J., et al. (2014) A multi-split mapping algorithm for circular RNA, splicing, trans-splicing and fusion detection. *Genome Biol*, 15.
30. Vromman,M., Vandesompele,J. and Volders,P.-J. (2021) Closing the circle: current state and perspectives of circular RNA databases. *Brief Bioinform*, 22, 288–297.

31. Vromman,M., Anckaert,J., Vandesomepele,J. and Volders,P.-J. (2022) CIRCprimerXL: Convenient and High-Throughput PCR Primer Design for Circular RNA Quantification. *Frontiers in Bioinformatics*, 2.
32. Zhang,J., Hou,L., Zuo,Z., Ji,P., Zhang,X., Xue,Y. and Zhao,F. (2021) Comprehensive profiling of circular RNAs with nanopore sequencing and CIRI-long. *Nat Biotechnol*, 39, 836–845.
33. Rahimi,K., Venø,M.T., Dupont,D.M. and Kjems,J. (2021) Nanopore sequencing of brain-derived full-length circRNAs reveals circRNA-specific exon usage, intron retention and microexons. *Nature Communications* 2021 12:1, 12, 1–15.
34. Xin,R., Gao,Y., Gao,Y., Wang,R., Kadash-Edmondson,K.E., Liu,B., Wang,Y., Lin,L. and Xing,Y. (2021) isoCirc catalogs full-length circular RNA isoforms in human transcriptomes. *Nature Communications* 2021 12:1, 12, 1–11.
35. Liu,Z., Tao,C., Li,S., Du,M., Bai,Y., Hu,X., Chen,J. and Yang,E. (2021) circFL-seq reveals full-length circular RNAs with rolling circular reverse transcription and nanopore sequencing. *Elife*, 10, e69457.

## **METHODS**

### **Study set-up**

As this study includes executing and evaluating circRNA detection tools, the co-authors can be divided into two groups 1) a first independent group (with no circRNA detection tool of their own) which initiated and designed the study and performed all the wet-lab work and data analysis (the validation co-author group), and 2) a second group of tool developer co-authors, which detected circRNAs using their own circRNA detection tools according to their expertise (the circRNA prediction co-author group) (details in Author Contribution section). During the study, meetings and emails were used to share the results (initially in a blinded manner) and discuss the final manuscript with the circRNA prediction co-authors.

### **Cell culture**

Three cancer cell lines from different cellular origin were randomly chosen as biological replicates. Ethical approval was obtained for this study (#EC014-202, Ghent University Hospital) and the cell lines were purchased from the JCRB Cell Bank (HLF and NCI-H23) or ECCAC (SW480). SW480 cells were cultured at 37 °C, 0% CO<sub>2</sub> in Leibovitz's L-15 medium (#31415-029, ThermoFisher). HLF cells and NCI-H23 cells were cultured at 37 °C, 5% CO<sub>2</sub> in DMEM, low glucose, GlutaMAX Supplement, pyruvate (#21,885,025, ThermoFisher) and RPMI 1640 Medium, HEPES (#52,400,041, ThermoFisher), respectively. 10% fetal bovine serum (FBS) (#F7524, Sigma) and 1% penicillin-streptomycin (10,000 U/mL) (#15,140,122, ThermoFisher) were added to all three media.

### **RNA isolation**

RNA was isolated from the cells using the miRNeasy Mini kit (#217004, Qiagen) according to the manufacturer's instructions, including the optional on-column DNase treatment (#79254, Qiagen). For each cell line, a sufficient number of cells was cultured to be able to harvest a minimum of 330 µg RNA. The RNA concentration was measured spectrophotometrically using a NanoDrop instrument and the RNA integrity was evaluated using a Fragment Analyzer. For each cell line, the RNA was pooled and aliquoted (1000 ng RNA in 100 µL nuclease-free water per aliquot) and stored at –80 °C, making a uniform RNA collection to use for all downstream experiments.

### **RNase R treatment, library preparation, and sequencing**

For each cell line, two aliquots of 1000 ng input RNA (in 10 µl nuclease-free water) were used. First, ribosomal RNA (rRNA) was removed with the NEBNext rRNA Depletion Kit (#E6350X,

New England Biolabs), following the manufacturer's instructions. Next, RNase R treatment was performed according to our previously described protocol (1). In summary, one aliquot of each cell line was treated with RNase R (#RNR07250 (250 U), Lucigen), and one aliquot of each cell line was treated as a buffer control. This was followed by a clean-up step using Vivacon 500, 10,000 MWCO Hydrosart columns (VN01H02, Sartorius). Subsequently, the NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (#E7760L, New England Biolabs) was used in combination with the NEBNext Multiplex Oligos for Illumina (#E7600S, New England Biolabs) to index and prepare the samples for sequencing. The library preparation protocol was adjusted to obtain relatively long insert sizes (average size of 636 nucleotides measured using a Fragment Analyzer): RNA fragmentation of 7.5 minutes; first-strand cDNA synthesis elongation step of 50 minutes instead of 15 minutes. The last bead clean-up step was performed twice to completely remove all indexes from the samples. Finally, the samples were pooled equimolarly and sequenced on a NovaSeq 6000 instrument using a NovaSeq 6000 S1 Reagent Kit v1.5 (300 cycles) (#20028317, Illumina), resulting in approximately 300 million paired-end 150-nucleotide reads per sample. Raw FASTQ files are stored in the Sequence Read Archive (PRJNA789110: SRX13414572 (untreated HLF), SRX13414573 (untreated NCI-H23), SRX13414574 (untreated SW480), SRX13414575 (RNase R treated HLF), SRX13414576 (RNase R treated NCI-H23), SRX13414577 (RNase R treated SW480)).

### **CircRNA detection**

In November 2020, a comprehensive list of all published circRNA detection tools was compiled, and all developers were invited to collaborate. Upon consent, they were asked to detect circRNAs using their own circRNA detection tool as they seemed fit for the data that was provided. The circRNA detection steps for each tool are detailed in the Supplementary Notes. Often, the default parameters were used as most of the methods included in our benchmarking underwent continuous development during the last several years and their parameters have been optimized for standard RNA sequencing data (as is the case in this study). We were unable to get into contact with the authors of find\_circ (2) and decided to run this tool ourselves, as it is one of the most frequently cited and broadly used circRNA detection tools. Unfortunately, other well-performing tools (according to (3, 4)), such as MapSplice (5), could not be included. More recent tools, such as Circall (6) and CYCLEr (7), have been published after the validation experiments were performed, and are therefore not included. After collecting all circRNA detection results, a uniform list of circRNAs defined by their BSJ position (chr, start, end, strand) and the BSJ count for each tool was compiled (Hg38, 0-based).

### **CircRNA selection and primer design**

Guided by a pilot experiment assessing circRNA RT-qPCR detectability in function of abundance and RNA input amount (Supplementary Data 5, Supplementary Figure 11), for each tool, 80 high-abundance circRNAs (with a BSJ count of at least 5), and 20 low-abundance circRNAs (with a BSJ count below 5) were selected (as two separate count bins). Primer pairs were designed using our primer design tool CIRCprimerXL (8). All primer sequences are available in Supplementary Table 3. If no primer pair could be designed for a given circRNA, a substitution was randomly selected from the complete dataset, considering the BSJ count bin. In total, 1,560 circRNA/tool/cell line tuples were selected. As some circRNAs were selected more than once (for different tools) the total number of unique circRNA/cell line pairs is 1,516, and the number of unique circRNAs (not taking into account the strand) is 1,457 (Supplementary Figure 14). Additionally, most of the selected circRNAs are detected by multiple tools (for which they were not selected). For the precision calculations, only the 20 + 80 selected circRNAs for a specific tool were used to evaluate that tool to keep the number of observations equal for each tool, even though more of its predicted circRNAs might have been validated. However, for the sensitivity calculations, the complete set of circRNAs had to be used (vide infra).

### **RNase R and RT-qPCR**

The RNA aliquots derived from the three cell lines were used for the circRNA RT-qPCR validation. A total of 1,080, 900, and 780  $\mu\text{L}$  RNA (100 ng/ $\mu\text{L}$ ) was required to validate 579, 500, and 437 circRNAs in HLF, NCI-H23, and SW480 cells, respectively. RNase R treatment was performed according to our previously reported protocol (1), adapted for this large-scale experiment. In summary, one RNA aliquot of a given cell line was treated with RNase R (#RNR07250 (250 U), Lucigen) and another was treated as a buffer control, for a total of 92 RNase R treated replicates and 92 buffer control replicates (2 \* 36 for HLF, 2 \* 30 for NCI-H23, and 2 \* 26 for SW480 RNA). All volumes were doubled during the buffer and RNase R reaction (total reaction volume of 20  $\mu\text{L}$ ). This was followed by a clean-up step using Vivacon 500, 10,000 MWCO Hydrosart columns (#VN01H02, Sartorius). Next, the RT reaction was performed on the 184 separate replicates using the iScript Advanced cDNA Synthesis Kit (#172-5038, Bio-Rad), according to the manufacturer's instructions. After RT, the cDNA was diluted 1:2 and an aliquot (2.5  $\mu\text{L}$ ) was further diluted 1:4 to evaluate the success of the RNase R reaction for each individual replicate. For this, ACTB and a known circRNA (chr1:117402185-117420649) previously described (1) (primer sequences available in Supplementary Table 10) were measured with qPCR using 2.5  $\mu\text{L}$  2x SsoAdvanced Universal SYBR Green Supermix (#172-5274, Bio-Rad), 0.5  $\mu\text{L}$  forward and reverse primer (5 nM), and 2  $\mu\text{L}$  cDNA per well, with qPCR duplicates. Once the RNase R treatment was successfully validated, all cDNA replicates were pooled per cell line and treatment condition. The cDNA was diluted 1:5 in 2x SsoAdvanced Universal SYBR Green Supermix (#172-5274, Bio-Rad). All 1,560 circRNA primer pairs were ordered from IDT in 96-well plates at a concentration of 100  $\mu\text{M}$  in nuclease-free water. All primers were diluted 1:160 to obtain a 0.625  $\mu\text{M}$  concentration. In each well of a qPCR plate, 2  $\mu\text{L}$  diluted primers and 3  $\mu\text{L}$  cDNA-master mix combination were added, resulting in an equivalent of 25 ng input RNA per qPCR reaction. Each assay (circRNA) was measured 4 times to include qPCR duplicates and to measure the abundance in both an RNase R untreated and treated sample, resulting in a total of more than 6000 qPCR reactions. A pipetting robot (EVO100, TECAN L) was used to dilute the primers and fill the qPCR plates. The qPCR reactions were run on a CFX384 instrument (Bio-Rad). Cq calling was done using the Bio-Rad CFX Manager (v3.1), with the 'regression' settings. The plates were stored at -20 °C prior to amplicon sequencing.

### **Amplicon sequencing**

After RT-qPCR, ~80% of the circRNAs were randomly included for amplicon sequencing. To make the sequencing library, the amplicons were pooled by combining 2  $\mu\text{L}$  of the PCR reaction from one of the untreated qPCR duplicates, per cell line. Next, the 3 samples were cleaned using Vivacon 500, 10,000 MWCO Hydrosart columns (#VN01H02, Sartorius). The PCR product pools were analyzed using a TapeStation 4150 (Agilent) and the concentration was measured using a Qubit fluorometer (ThermoFisher). Next, the three pools were diluted in nuclease-free water to obtain 50  $\mu\text{L}$  samples with a concentration of 20 ng/ $\mu\text{L}$ . Finally, the samples were prepared for sequencing using the NEBNext Ultra II DNA Library Prep Kit for Illumina (#E7645S, New England Biolabs) and NEBNext Multiplex Oligos for Illumina (Dual Index Primers Set 1) (#E7600S, New England Biolabs). To retain all amplicons, no size selection was performed after adaptor ligation, and 1.0x AMPure XP beads (#A63881, Beckman Coulter) in a 1:1 sample:beads ratio was used instead. After library preparation, the samples were pooled equimolarly. The pool was sequenced on a NextSeq 500 instrument using a Mid Output Kit v2.5 (150 cycles) (#20024904, Illumina), resulting in approximately 25-30 million paired-end 75-nucleotides reads per library.

### **Data analysis**

Data analysis was mostly done using R (9) (v4.2.1) in RStudio (10) (v2022.07.1). The following R packages were used: tidyverse (v1.3.2), conflicted (v1.1.0), ggrepel (v0.9.1), ggseqlogo (v0.1), europepmc (v0.4.1), gplots (v3.1.3), ggpubr (v0.4.0), quantreg (v5.94), rstatix (v0.7.0) and UpSetR (v1.4.0). For sequencing data analyses, including circRNA detection and amplicon sequencing analysis, the Ghent University high-performance cluster was used. For

this, Python3 (v3.6.8) (11), Bowtie2 (v2.3.4.1) (12), fastahack (v1.0.0), SAMtools (v1.11) (13), and BEDTools (v2.30.0) (14) were used. The human reference transcriptome was downloaded as a GTF file from Ensembl (15). All data analysis scripts are available at [https://github.com/OncoRNALab/circRNA\\_benchmarking](https://github.com/OncoRNALab/circRNA_benchmarking).

#### Amplicon sequencing data analysis

For the amplicon sequencing data analysis, first, a custom Python script matches the primer sequences with the first 16-mer of each read (forward and reversed) and generates a separate FASTQ file per primer pair, containing all reads starting with that primer sequence. The FASTQ reads are then clipped to remove the primer sequences. Next, all FASTQ files are mapped against the reference genome (Ensembl GRCh38.101) supplemented with the theoretical BSJ amplicon sequences using Bowtie2 with default settings. Lastly, the Bowtie2 BAM files are converted to counts using another custom Python script and the percentage on-target amplification was calculated for each primer pair.

#### Determination of orthogonal precision values and sensitivity

Several strategies to filter the data prior to precision and sensitivity calculations were explored. For RT-qPCR, a circRNA was considered validated when at least one of the untreated RNA samples had a Cq value above 10. Multiple variations of this threshold and a potential upper Cq threshold were evaluated. For RNase R validation, a subset of circRNAs with at least one untreated replicate with a Cq value below 32 was selected to ensure that the enzymatic degradation of a false positive circRNA could be measured. A circRNA was considered validated upon RNase R treatment if the difference in Cq between the untreated and treated RNA sample was equal to or less than 3 cycles, based on a previous study (1). As there were two qPCR replicates available for each (un)treated sample, the 'best-case scenario' was used to calculate the difference in Cq by subtracting the maximum untreated Cq replicate from the minimum treated Cq replicate. A circRNA with both untreated replicates having a Cq value above 32 was labeled as NA. For amplicon sequencing, a circRNA was considered validated if the primer pair was found in at least 1000 reads and if at least 50% of these reads matched the expected amplicon upon mapping with Bowtie2. For a random subset of circRNAs, unintentionally no amplicon sequencing was performed; these were labeled as NA. A detailed description of the choice of performance metrics is available in Supplementary Data 12 and 13. To calculate precision values per tool, BSJ count bin, and validation method, the number of circRNAs that passed the validation was divided by the total number of circRNAs that were not NA for that validation method. We also determined a compound precision value by considering both qPCR, RNase R treatment, and amplicon sequencing. For this, each circRNA was labeled as a true positive (i.e., validated by all three methods), as a false positive (i.e., not validated by at least one of the methods), or as NA (i.e., not included in the amplicon sequencing run). Based on this summarizing label, compound precision values were computed for each tool and BSJ count bin. The number of theoretically true positive circRNAs was calculated by multiplying the total number of circRNAs predicted by that tool for that sample with the compound precision value (i.e., the extrapolated sensitivity). The sensitivity was also calculated as the percentage of circRNAs each tool detected from the validated set of true positive circRNAs (i.e., the circRNAs labeled as true positives over all three methods). This metric should be used with caution as it is based on a biased selection of circRNAs due to the overlap among tools (Supplementary Data 12). To calculate the sensitivity per BSJ count group, the median BSJ count of each circRNA was used (as most circRNAs are detected by multiple tools and have therefore multiple BSJ count values).

#### Annotation of circRNAs

To obtain the circRNA splice site information, the BSJ-flanking nucleotides were extracted from the reference genome using fastahack (Ensembl GRCh38.104). To compare BSJ positions with known linear annotation, BEDtools intersect was used with a list of canonical transcripts from Ensembl with their positions based on the corresponding Ensembl GTF file (Ensembl GRCh38.103). When a circRNA mapped to multiple isoforms, the annotation was

labeled as 'ambiguous' and the circRNA was not taken into account for further annotation-based calculations and figures. The annotation was used to compute the length of each circRNA excluding introns, and the number of exons per circRNA. CircRNAs smaller than their host gene exon were labeled 'single-exon' circRNAs. For the length of each circRNA including introns, the BSJ start position was simply subtracted from the BSJ end position. Furthermore, for each circRNA, annotation was added to indicate if the BSJ start and end positions match known exon boundaries. When comparing predicted circRNAs to circRNAs previously described in databases, strand information was discarded.

### Combination of tools

To compare the circRNA tools, the union and intersection of all circRNAs predicted by each tool pair and triple were calculated. A weighted precision value was calculated for each combination of tools as follows:  $((\text{compound\_precision\_1} * \text{total\_n\_1}) + (\text{compound\_precision\_2} * \text{total\_n\_2})) / (\text{total\_n\_1} + \text{total\_n\_2})$ . For this, strand information was discarded, as 4 out of 16 tools did not report circRNA strands and would therefore have been excluded. These calculations were performed for each cell line separately. To determine the correspondence among tools, the Jaccard distance was calculated and heatmap clusters were generated. The tools were compared based on the mere presence or absence of a circRNA. Also, for the calculation of how many tools detected a given circRNA, circRNA strand information was discarded.

### Statistical analyses

To evaluate the effect of circRNA characteristics on circRNA validation, the Chi-squared test was used (`chisq.test()` function in R). For every test, the set of used circRNAs was slightly different depending on the availability of annotation information. All tests had all expected values in the contingency table above 5, therefore no correction for small sample size was necessary. Seven different characteristics were tested, and no multiple testing correction was performed. To evaluate the effect of circRNA detection tool methods on sensitivity and to evaluate the effect of the combination of tools with different approaches, the two-sided Mann-Whitney U test was used (`rstatix::wilcox_test()` function in R). For correlation analysis between the sensitivity and the extrapolated sensitivity, the Spearman rank correlation was used (`cor.test(method = 'spearman')` function in R). For correlation analysis between circRNA BSJ counts from different tools, or between circRNA BSJ counts and Cq values, or between Cq values in different cell lines, linear models were used (`lm()` function in R). To evaluate the contribution of the cell lines (in contrast to the tools) to the precisions and sensitivity values, an ANOVA test was used (`aov()` function in R).

## **DATA AVAILABILITY**

We anticipate this study will serve as a future resource for the circRNA community. The information on all predicted circRNAs ( $n = 315,312$ ), including the large extensively validated circRNA set ( $n = 1,516$ ), along with the validation results are available in the GitHub repository [https://github.com/OncoRNALab/circRNA\\_benchmarking](https://github.com/OncoRNALab/circRNA_benchmarking) and as Supplementary Tables. The set of circRNAs previously described in databases (Circ2Disease, circad, CircAtlas, circbank, circBase, CIRCpediav2, CircR2disease, CircRiC, circRNADb, CSCD, exoRBase, MiOncoCirc, and TSCD) is also included in the GitHub repository. All databases were accessed in the context of a previous study (30). Raw FASTQ files are stored in the Sequence Read Archive (PRJNA789110: SRX13414572 (untreated HLF), SRX13414573 (untreated NCI-H23), SRX13414574 (untreated SW480), SRX13414575 (RNase R treated HLF), SRX13414576 (RNase R treated NCI-H23), SRX13414577 (RNase R treated SW480)).

## **CODE AVAILABILITY**

All the scripts used to compute the metrics described in the study and generate the figures are available at [https://github.com/OncoRNALab/circRNA\\_benchmarking](https://github.com/OncoRNALab/circRNA_benchmarking).

## METHODS-ONLY REFERENCES

1. Vromman,M., Yigit,N., Verniers,K., Lefever,S., Vandesompele,J. and Volders,P. (2021) Validation of Circular RNAs Using RT-qPCR After Effective Removal of Linear RNAs by Ribonuclease R. *Curr Protoc*, 1, e181.
2. Memczak,S., Jens,M., Elefsinioti,A., Torti,F., Krueger,J., Rybak,A., Maier,L., Mackowiak,S.D., Gregersen,L.H., Munschauer,M., et al. (2013) Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature*, 495, 333–338.
3. Hansen,T.B., Venø,M.T., Damgaard,C.K. and Kjems,J. (2015) Comparison of circular RNA prediction tools. *Nucleic Acids Res*, 44, e58.
4. Zeng,X., Lin,W., Guo,M. and Zou,Q. (2017) A comprehensive overview and evaluation of circular RNA detection tools. *PLoS Comput Biol*, 13, e1005420.
5. Wang,K., Singh,D., Zeng,Z., Coleman,S.J., Huang,Y., Savich,G.L., He,X., Mieczkowski,P., Grimm,S.A., Perou,C.M., et al. (2010) MapSplice: Accurate mapping of RNA-seq reads for splice junction discovery. *Nucleic Acids Res*, 38, 1–14.
6. Nguyen,D.T., Trac,Q.T., Nguyen,T.H., Nguyen,H.N., Ohad,N., Pawitan,Y. and Vu,T.N. (2021) Circall: fast and accurate methodology for discovery of circular RNAs from paired-end RNA-sequencing data. *BMC Bioinformatics*, 22.
7. Stefanov,S.R. and Meyer,I.M. (2022) CYCLeR—a novel tool for the full isoform assembly and quantification of circRNAs. *Nucleic Acids Res*, 1, 13–14.
8. Vromman,M., Anckaert,J., Vandesompele,J. and Volders,P.-J. (2022) CIRCprimerXL: Convenient and High-Throughput PCR Primer Design for Circular RNA Quantification. *Frontiers in Bioinformatics*, 2.
9. R Core Team (2019) R: A language and environment for statistical computing.
10. RStudio Team (2020) RStudio: Integrated Development for R. RStudio.
11. van Rossum,G. and Drake,F.L. (2009) Python 3 Reference Manual.
12. Langmead,B. and Salzberg,S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nature Methods* 2012 9:4, 9, 357–359.
13. Danecek,P., Bonfield,J.K., Liddle,J., Marshall,J., Ohan,V., Pollard,M.O., Whitwham,A., Keane,T., McCarthy,S.A., Davies,R.M., et al. (2021) Twelve years of SAMtools and BCFtools. *Gigascience*, 10.
14. Quinlan,A.R. and Hall,I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26, 841–842.
15. Cunningham,F., Allen,J.E., Allen,J., Alvarez-Jarreta,J., Amode,M.R., Armean,I.M., Austine-Orimoloye,O., Azov,A.G., Barnes,I., Bennett,R., et al. (2022) Ensembl 2022. *Nucleic Acids Res*, 50, D988–D995.