

# Comparative Performance of the Genolab M and Novaseq 6000 Sequencing Platforms for Transcriptome and LncRNA Analysis

**Yongfeng Liu**

GeneMind Biosciences Company Limited

**Ran Han**

Beijing Guoke Biotechnology co., LTD

**Letian Zhou**

GeneMind Biosciences Company Limited

**Mingjie Luo**

GeneMind Biosciences Company Limited

**Lidong Zeng**

GeneMind Biosciences Company Limited

**Xiaochao Zhao**

GeneMind Biosciences Company Limited

**Yukun Ma**

Beijing Guoke Biotechnology co., LTD

**Zhiliang Zhou**

GeneMind Biosciences Company Limited

**Lei Sun** (✉ [sunlei@genemind.com](mailto:sunlei@genemind.com))

GeneMind Biosciences Company Limited

---

## Research Article

**Keywords:** GenoLab M, NovaSeq 6000, transcriptome, LncRNA, compatibility

**Posted Date:** September 20th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-900102/v1>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

**Version of Record:** A version of this preprint was published at BMC Genomics on November 17th, 2021.  
See the published version at <https://doi.org/10.1186/s12864-021-08150-8>.

# Abstract

**Background:** GenoLab M is a recently established next-generation sequencing platform from GeneMind Biosciences. Presently, Illumina sequencers are the globally leading sequencing platform in the next-generation sequencing market. Here, we present the first report to compare the transcriptome and LncRNA sequencing data of the GenoLab M sequencer to NovaSeq 6000 platform in various types of analysis.

**Results:** We tested 16 libraries in three species using various library kits from different companies. We compared the data quality, genes expression, alternatively spliced (AS) events, single nucleotide polymorphism (SNP), and insertions–deletions (InDel) between two sequencing platforms. The data suggested that platforms have comparable sensitivity and accuracy in terms of quantification of gene expression levels with technical compatibility.

**Conclusions:** Genolab M is a promising sequencing platform for transcriptomics and LncRNA studies with high performance at low costs.

## Background

The past dozens of years have witnessed a new era in functional genomics using sequencing technologies[1]. The launch of the Roche 454 sequencer opened the era of next-generation sequencing (NGS)[2]. Compared with the traditional Sanger sequencing technology[3], NGS has significantly higher throughput and reduced costs [1]. Taking advantages of the power of NGS, transcriptome and Long non-coding RNA (LncRNA) sequencing has been accepted as a mainstream profiling technique to reveal gene regulatory networks in both animals and plants[4].

In the short history of NGS era, many sequencing platforms have emerged: Roche 454, Illumina series (GA, HiSeq, NextSeq, NovaSeq, etc.)[5], BGI (BGISEQ-500)[6], Ion Torrent[7], GenapSys[8]. These platforms employ different sequencing chemistry and detection approaches, and each of them has specific advantages and shortcomings[9]. After years of technology evolution and product commercialization, Illumina sequencers become the most widely used platform. However, the high instrument and reagent cost hinders broader applications[10]. In recent years, BGI's MGI sequencers have received more attention in their cost effectiveness[11], though BGI's unique DNB (DNA Nanoball) sequencing approach requires complicated library preparation and quality control procedure[12–15]. As DNA sequencing applications increase in different research fields and clinical settings, there is still a need to develop sequencers that are accurate, flexible, and cost-efficient for applications.

Recently, GeneMind Biosciences Company Limited (GeneMind), launched a new sequencing instrument (GenoLab M™) based on their previous work on GenoCare™ single molecule sequencer[16]. An overview of the mechanism of GenoLab M DNA sequencer is outlined in Fig. 1. The GenoLab M sequencer employs sequencing-by-synthesis techniques and applies reversible termination approaches. Instead of single molecule detection, surface-based amplification is performed after target DNA capture to enhance signal-to-noise ratio of sequencing. After amplification, the amplified DNA colonies on the flow cell are

hybridized to a sequencing primer, which contains a target-complimentary sequence. Fluorescence-dye labeled nucleotides and a polymerase are applied to start sequencing. Four-color fluorescence signals are collected cycle-by-cycle to detect the incorporated nucleotides.

Previously, GenoLab M's performance on transcriptome and LncRNA has not yet been evaluated by the scientific community. Here, we characterized the performance of GenoLab M on transcriptome and LncRNA by parallel comparison with NovaSeq 6000 on three different species: mouse, bean, and human. The raw data quality, gene expression level, alternatively spliced (AS) events, single nucleotide polymorphism (SNP), and insertions–deletions (InDel) analysis from the two sequencing platforms are compared. The data suggest that the GenoLab M is a promising sequencing platform for transcriptomics and LncRNA studies in animal, plant and human with comparable performance at lower cost.

## Method

### Samples preparation and RNA extraction

Mouse testicular tissue, human Lieming Xu-2 cells and bean hairy root tissue were collected for RNA extraction. RNAs were extracted by HiPure Universal RNA Mini Kit (Guangzhou Magen Biotechnology Co., Ltd.). Total RNA concentration and purity and integrity were measured via NanoDrop 2000 (Thermo Fisher Scientific, Wilmington, DE) and RNA Nano 6000 Assay Kit of the Agilent Bioanalyzer 2100 system (Agilent Technologies, CA, USA), respectively.

### Transcriptome and LncRNA sequencing

Transcriptome library construction were performed by Hieff NGS Ultima Dual-mode mRNA Library Prep Kit for Illumina (Yeasen Biotechnology (Shanghai) Co., Ltd. China), Fast RNA-seq Lib Prep Module for Illumina (ABclonal Technology Co.,Ltd, China), TIANSeq Stranded RNA-Seq Kit (Illumina) (TIANGEN Biotech (Beijing) Co., Ltd, China) and VAHTS Universal V6 RNA-seq Library Prep Kit for Illumina (Vazyme Biotech Co., Ltd, China). These mRNA libraries were marked as Mouse, Human or Bean\_mRNA\_YS, AB, TG or VZ. LncRNA library were constructed via Hieff NGS Ultima Dual-mode RNA Library Prep Kit for Illumina and Hieff NGS MaxUp rRNA Depletion Kit(human/mouse/rat) (Yeasen Biotechnology (Shanghai) Co., Ltd. China), VAHTS Universal V6 RNA-seq Library Prep Kit for Illumina and Ribo-off rRNA Depletion Kit (Human/Mouse/Rat)(Vazyme Biotech Co.,Ltd, China), TIANSeq Stranded RNA-Seq Kit (Illumina) and TIANSeq rRNA Depletion Kit (H/M/R) (NR101-TA) (TIANGEN Biotech (Beijing) Co.,Ltd,China). These libraries were marked as Mouse or Human\_LncRNA\_YS, VZ or TG. After library QC, they were subjected to NovaSeq 6000 and GenoLab M sequencing in PE150 or PE100 mode.

### Cross-platform mRNA and LncRNA sequencing data analysis

Raw sequencing reads in fastq format were processed through a GeneMind in-house perl pipeline. Reads containing adapter, ploy-N or low quality reads were filtered out to get clean reads. These clean reads were then mapped to the reference genome sequence with a perfect match or one mismatch method via

HISAT2 tools software[17]. The corresponding genome references were downloaded from ensemble database by [ftp://ftp.ensembl.org/pub/release-101/fasta/homo\\_sapiens/dna/](ftp://ftp.ensembl.org/pub/release-101/fasta/homo_sapiens/dna/), [ftp://ftp.ensembl.org/pub/release-101/fasta/mus\\_musculus/dna](ftp://ftp.ensembl.org/pub/release-101/fasta/mus_musculus/dna/), and [ftp://ftp.ensemblgenomes.org/pub/plants/release-48/fasta/glycine\\_max/dna/](ftp://ftp.ensemblgenomes.org/pub/plants/release-48/fasta/glycine_max/dna/). StringTie [18] was then used for transcript reconstruction. As candidate genes were defined as genes which were mapping to unannotated transcribed region, meanwhile, coding peptide was more than 50 amino acid residues with two or more exons. SNP and InDel calling was carried out by using GATK [19], furthermore, SnpEff [20] was used to annotate these mutations. Raw vcf files were filtered with GATK standard filter method and other parameters (clusterWindowSize:10; MQ0 >= 4 and (MQ0/(1.0\*DP)) > 0.1; QUAL < 10; QUAL < 30.0 or QD <5.0 or HRun > 5), and only SNPs with distance > 5 were retained. Alternative spliced events were identified by ASprofle software [21]. Expression values of candidate genes (FPKM) were calculated by RSEM [22].

For LncRNA identification, bioinformatic pipeline was performed according to published methods [23] with minor modifications. The transcriptome was assembled using the StringTie based on the reads mapped to the reference genome. The known LncRNAs from the assembled transcripts are defined using the Cuffcompare program from the Cufflinks package. The remaining transcripts (unknown transcripts) were used to screen for putative LncRNAs. Transcripts of more than 200 nt length and two exons were selected as candidate LncRNA transcripts. Then, CPC [24], CNCI [25], Pfam [26] and CPAT [27] were used to distinguish the protein-coding genes from the non-coding genes, and inter set as putative LncRNA. As well as the different types of LncRNAs including lincRNA, intronic LncRNA, anti-sense LncRNA, sense LncRNA were selected using gffcompare. StringTie (1.3.1) [28] was used to calculate FPKMs of LncRNAs. The FPKM of novel LncRNAs must be  $\geq 0.1$ .

## Results

### Base and raw data quality

Following RNA extraction, two aliquots of each extract were constructed as Illumina libraries, respectively, using identical amounts of starting material, and then subsequently sequenced to facilitate bioinformatic comparisons on the data. In addition, to verify the compatibility of the library preparation kit for GenoLab M, we used kits from different manufacturers for testing (Supplemental Table S1). The sequencing strategy was pair-end 100 bp for GenoLab M and pair-end 150 bp for NovaSeq 6000. We initially generated between 23.20 M to 62.87 M clean reads per library in NovaSeq 6000 platform, and 26.86 M to 139.69 M clean reads per library in GenoLab M platform (Table 1). Each individual sample has similar base throughput from both sequencing platforms. The quality of sequencing data was checked using FastQC. For high base quality (over Q20) base percentages, the GenoLab M showed an average of 94.86%, and the NovaSeq 6000 showed an average of 97.50% with a slight preponderance (Table 1). As shown in Fig. 2, the clean reads from GenoLab M reached an average mapping rate of 91.80% and an average unique mapping rate of 88.33%, which are comparable to the mapping rates of reads from the NovaSeq 6000 platform. The two platforms shared fairly consistent reads distribution along genes

across species (Fig.3) and in expression density distribution (Fig.4). Interestingly, the LncRNA expression level measured using Yeasen LncRNA library kit (YS) is higher than the other kits used in human and mouse. In Fig.5, the charts showed that accuracy in the quantification of both low and high abundance genes were consistent. They further indicate that LncRNA expression by YS has obviously higher abundance than the other kits in human and mouse (Fig.5 A and B), which is consistent with the Fig.4 B and D. Overall, the sequence quality of the two platforms was similar across various library kits.

### **Inter-platform comparison of gene detection and quantification**

In transcriptome and LncRNA analysis, the identification of genes is very important for the majority of research projects. Therefore, we further compared the capacity of GenoLab M and NovaSeq 6000 platforms on gene detection and quantification. Totally over 42,000, 16,000 and 26,000 genes were identified in bean, human, and mouse, respectively, via two sequencing platform (Fig.6, A-C). For transcriptome, we observed a small fraction of different genes between the GenoLab M and NovaSeq 6000 platforms. Over 92% of genes were commonly detected by both sequencing platforms. However, for LncRNA, only 71% of genes were shared between the two sequencing platforms (Fig.6, D-E). This difference most likely stemmed from analysis using the method StringTie as novel LncRNAs judgment and the different read length of the sequence [18]. StringTie (1.3.1) was used to calculate FPKMs of LncRNAs and novel LncRNA was set at least 0.1. We checked the Pearson correlation coefficient of the transcriptome and LncRNA data produced by the two platforms using the same methods and found that all one pairs of samples showed high correlation coefficients, ranging from 0.972 to 0.992 in transcriptome, and ranging from 0.691 to 0.793 in LncRNA (Fig. 7). There is still a slight gap in the correlation between LncRNA and the two platforms. In all, GenoLab M has remarkable inter-platform concordance with NovaSeq 6000, suggesting that GenoLab M could substitute NovaSeq 6000 in many application fields where transcriptome and LncRNA are the primary focus.

### **Detection of alternative splicing**

As one of the major mechanisms to generate transcriptome diversity, alternative splicing (AS) is gaining more and more attention in recent years. In this context, the ability of each sequencing platform under comparison to detect splicing junctions and corresponding alternative splicing patterns were subsequently analyzed across transcriptomes. In mouse, 53,557, 59,709 and 53,014, 56,741, 64,105 and 48,089 AS events could be detected by GenoLab M and NovaSeq 6000, respectively. Top three AS events in all libraries were TSS: Alternative 5' first exon (transcription start site), TTS: Alternative 3' last exon (transcription terminal site) and AE: Alternative exon ends (5', 3', or both) cross two platforms (Fig.8 A). In mouse LncRNA data, the AS events component in mRNA presented similarly to transcriptome (Fig.8 B). For human sample, AS events component in transcriptome and mRNA of LncRNA data were of the same pattern and Top 3 AS were TSS, TTS and SKIP:Skipped exon(SKIP\_ON,SKIP\_OFF pair) as showed in Fig.8 C and D. In beans, 78,137, 82,558 and 105,038, 83,072, 84,526 and 90,580 AS events could be detected by GenoLab M and NovaSeq 6000, respectively. Top three AS events in all libraries were TSS, TTS and AE

(Fig.8 E). As for both the number and the type of different AS events, we found that there was no significant difference between the three species in the two platforms.

### **Identification of SNP and InDel mutation**

SNP and InDel are crucial genomic features to reveal genetic variation. High throughput transcriptome analysis contributes to how these DNA variations can be transcribed into RNA messengers to affect subsequent protein function. Therefore, we examined the competency of the GenoLab M sequencing platform to detect SNP and InDel variations at the mRNA level. Regarding SNP detection, we found that SNPs called from the two sequencing platforms (Table.2) were highly similar in both variety and quantity. The largest difference is that the GenoLab M platform identified slightly more SNP events in mice than NovaSeq 6000 on average.

For InDel events, GenoLab M detected less of them than the NovaSeq 6000 in bean, human and mouse (Table.3). The closest InDel number was in bean sample prepared with Vazyme Biotech (VZ) transcriptome library kit, while significant difference was observed in mouse via Yeasen Biotechnology (YS) transcriptome library kit. These results suggest that GenoLab M has slightly inferior in InDel detection, probably due to shorter read length in this study.

## **Discussion**

With the advantages of high-throughput and low cost, NGS is becoming a powerful tool for scientific and clinical research. Increased sequencing accessibility and flexibility have not only broadened NGS applications, but also led to the development of novel sequencing platforms and sequencing methods in turn[29]. Currently, Illumina's sequencers are the globally leading sequencing platform. The NovaSeq 6000, its most powerful instrument, has prominent properties of lower error rate and less variation compared to other sequencers in the Illumina series[30]. It is able to generate 6 TB of sequencing data in a single run with a running cost between 12–18 USD/Gb[31]. GenoLab M, the new sequencer of GeneMind, can generate 300 Gb of sequence data in a single run with price per Gb cost less than half of that. In this study, we generated large transcript and LncRNA datasets from the two sequencing platforms across three model species (human: 4 mRNA, 3 LncRNA; mouse: 3 mRNA, 3 LncRNA; bean: 3 mRNA). Next, we compared the datasets obtained from the two platforms. To make our study as comprehensive as possible, we compared the quality of data, distribution of reads, gene expression, AS, SNP and InDel of the two platforms.

Our analysis of the data generated from two platforms showed that sequences from both instruments were of comparable quality with the exception that NovaSeq 6000 reads showed slightly higher Q20 percentage than GenoLab M. We are confident that higher quality data from GenoLab M are attainable through instrument hardware, software and reagent kit updates, given that the instrument was launched just last year[32, 33].

Gene expression has always been an important part of the research on transcriptome and LncRNA[33]. In the comparison of transcriptome genes expression, we found that there was no significant quality difference between the two platforms, and the correlation analysis showed high consistency. This indicates that GenoLab M can achieve a similar level of mRNA detection as NovaSeq 6000 and is suitable for use with the same library kits designed for Illumina sequencing. This compatibility enables users to test the sequencing platform with minimum transition cost and generate high quality sequencing data. We believe that this would make transcriptome sequencing more accessible for researchers. However, in the LncRNA area, we found that GenoLab M's performance had a small gap compared to NovaSeq 6000 in terms of gene expression correlation. We think sequencing read length may impact the LncRNA detection rate[34]. We plan to conduct further laboratory tests to determine the cause of this difference.

Our experimental results proved that GenoLab M could obtain equivalent data quality as NovaSeq 6000, in both mRNA and LncRNA level with 7 library preparation kits from 4 companies. This suggests that GenoLab M can be a viable substitute for NovaSeq 6000 in the RNA sequencing. However, our study does lack biological repeats, which could be supplemented in further work. We also realize that we still need to increase the number of samples and species to further demonstrate the reliability of the GenoLab M platform. In the future, we plan to work with more researchers in broader application areas to verify the capability and stability of the platform.

## Conclusions

In summary, we highlight that both GenoLab M and NovaSeq 6000 sequencing platforms have similar and comparable performance metrics (sensitivity and accuracy) and can capture genes, AS, and SNP at transcriptome and LncRNA levels. The GenoLab M offers a cost-effective alternative to the NovaSeq 6000 platform with similar data quality.

## Declarations

### Ethics approval and consent to participate

This study was approved by the Ethics Committee of GeneMind Biosciences Company Limited. All methods were carried out in accordance with relevant guidelines and regulations. Mouse testicular tissue, human Lieming Xu-2 cells and bean hairy root tissue were from Beijing Guoke Biotechnology co., LTD.

### Consent for publication

Not applicable.

### Availability of data and materials

The transcriptome and LncRNA data are available at SRA database with SUB10177917 and SUB10176628.

### Competing interests

The authors declare that they have no competing interests.

### Funding

Not applicable.

### Authors' contributions

Lei Sun conceived and designed the research, reviewed and revised the manuscript. Yongfeng Liu and XiaoChao Zhao wrote the manuscript. Zhiliang Zhou reviewed and revised the manuscript. Ran Han and Yukun Ma performed sample preparation and RNA extract, and preliminary analysis. Mingjie Luo took charge of sequencing. Letian Zhou supported data mining and figure drawing. Lidong Zeng developed GenoLab M's bioinformatics pipeline. All authors read and approved the final version of the manuscript.

### Acknowledgements

We would like to thank all current and past members of the GeneMind team who contributed to the development of the sequencing technology.

## References

1. Zheng J, Zhang H, Banerjee S, Li Y, Zhou J, Yang Q, Tan X, Han P, Fu Q, Cui X: **A comprehensive assessment of Next-Generation Sequencing variants validation using a secondary technology.** *Molecular genetics & genomic medicine* 2019, **7**(7):e00748.
2. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen Y-J, Chen Z: **Genome sequencing in microfabricated high-density picolitre reactors.** *Nature* 2005, **437**(7057):376–380.
3. Sanger F, Nicklen S, Coulson AR: **DNA sequencing with chain-terminating inhibitors.** *Proceedings of the National Academy of Sciences of the United States of America* 1977, **74**(12):5463–5467.
4. Xu Y, Lin Z, Tang C, Tang Y, Gao Q: **A new massively parallel nanoball sequencing platform for whole exome research.** *BMC Bioinformatics* 2018.
5. Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, Hall KP, Evers DJ, Barnes CL, Bignell HR: **Accurate whole human genome sequencing using reversible terminator chemistry.** *Nature* 2008, **456**(7218):53–59.
6. Huang J, Liang X, Xuan Y, Geng C, Li Y, Lu H, Qu S, Mei X, Chen H, Yu T: **A reference human genome dataset of the BGISEQ-500 sequencer.** *Gigascience* 2017, **6**(5):gix024.

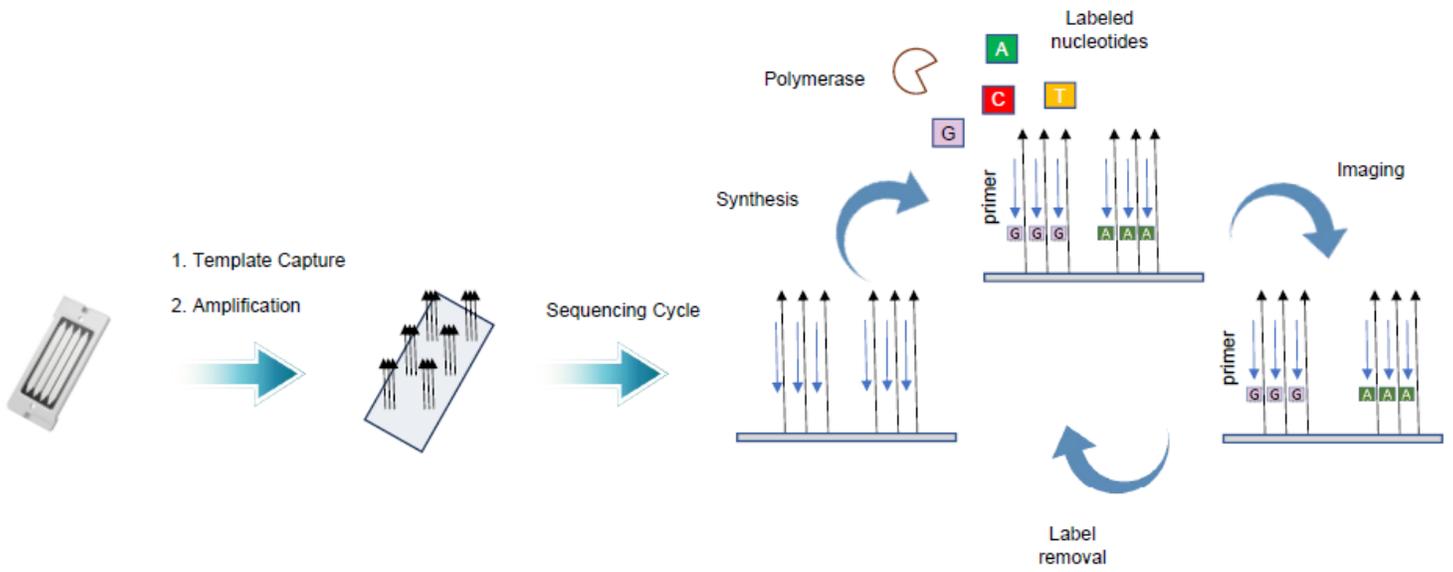
7. Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, Davey M, Leamon JH, Johnson K, Milgrew MJ, Edwards M: **An integrated semiconductor device enabling non-optical genome sequencing.** *Nature* 2011, **475**(7356):348–352.
8. Esfandyarpour H, Parizi KB, Barmi MR, Rategh H, Witney FR: **High accuracy DNA sequencing on a small, scalable platform via electrical detection of single base incorporations.** 2019.
9. Goodwin S, McPherson JD, McCombie WR: **Coming of age: ten years of next-generation sequencing technologies.** *Nature Reviews Genetics* 2016, **17**(6):333–351.
10. A., S.: **ILLUMINA TO SEQUENCE 100,000 U.K. GENOMES.** *Chemical & Engineering News* 2014, **92**(32):11–11.
11. Huang J, Liang X, Xuan Y, Geng C, Li Y, Lu H, Qu S, Mei X, Chen H, Yu T: **A reference human genome dataset of the BGISEQ-500 sequencer.** *Gigascience* 2017(5):1–9.
12. Zhu FY, Chen MX, Ye NH, Qiao WM, Bei G, Wai-Ki L, Yuan T, Dong Z, Di Z, Liu TY: **Comparative performance of the BGISEQ-500 and Illumina HiSeq4000 sequencing platforms for transcriptome analysis in plants.** *Plant Methods* 2018, **14**(1):69-.
13. Sandoval-Velasco M, Rodríguez J, Perezestrada C, Zhang G, Smith O: **Hi-C chromosome conformation capture sequencing of avian genomes using the BGISEQ-500 platform.** *GigaScience* 2020, **9**(8).
14. Hak-Min K, Sungwon J, Oksung C, Hoon JJ, Hui-Su K, Asta B, Hwang-Yeol L, Yu Y, Sung CY, Bolser DM: **Comparative analysis of 7 short-read sequencing platforms using the Korean Reference Genome: MGI and Illumina sequencing benchmark for whole-genome sequencing.** *GigaScience* 2021(3):3.
15. Wang Y, Dai Z, Zhang Z, Zhu L, Jiang L: **Draft genome sequence of a multidrug-resistant *Stenotrophomonas* sp. B1-1 strain isolated from radiation-polluted soil and its pathogenic potential.** *Journal of Global Antimicrobial Resistance* 2020.
16. Zhao L, Deng L, Li G: **Single molecule sequencing of the M13 virus genome without amplification.** *Plos One* 2017, **12**(12):34209.
17. Kim D, Langmead B, Salzberg SL: **HISAT: A fast spliced aligner with low memory requirements.** *Nature Methods* 2015, **12**(4).
18. Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL: **StringTie enables improved reconstruction of a transcriptome from RNA-seq reads.** *Nature Biotechnology* 2015, **33**(3):290–295.
19. Mckenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernysky A, Garimella K, Altshuler D, Gabriel S, Daly M: **The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data.** *Genome Research* 2010, **20**(9):1297–1303.
20. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM: **A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff.** *Fly* 2012, **6**(2):80–92.
21. Florea L, Song L, Salzberg SL: **Thousands of exon skipping events differentiate among splicing patterns in sixteen human tissues.** *F1000 Research* 2013, **2**:188.

22. Dewey CN, Li B: **RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome.** *BMC Bioinformatics* 2011, **12**(1):323–323.
23. Luo H, Bu D, Sun L, Fang S, Liu Z, Yi Z: **Identification and function annotation of long intervening noncoding RNAs.** *Briefings in Bioinformatics* 2017.
24. Sun L, Luo H, Bu D, Zhao G, Yu K, Zhang C, Liu Y, Chen R, Zhao Y: **Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts.** *Nucleic Acids Research* 2013(17):e166-e166.
25. Kai C, Lei W, Xu C, Zhang X, Xing L, Zuo H, Zhou Z, Jia P, Liu S, Qu G: **Influence on surgical treatment of intertrochanteric fracture with or without preoperative skeletal traction.** *Chinese Journal of Joint Surgery(Electronic Edition)* 2015, **41**(6):e74-e74.
26. Bateman, A.: **The Pfam Protein Families Database.** *Nucleic Acids Research* 2000, **28**(1):263–266.
27. Chen G, Wang Z, Wang D, Qiu C, Liu M, Xing C, Zhang Q, Yan G, Cui Q: **LncRNADisease: a database for long-non-coding RNA-associated diseases.** *Nucleic Acids Research* 2013(D1):D983-D986.
28. Kim, Daehwan, Leek, Jeffrey, T., Pertea, Geo, M., Salzberg, Steven: **Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown.** *Nature Protocols Erecipes for Researchers* 2016.
29. Mardis ER: **The impact of next-generation sequencing technology on genetics.** *Trends in Genetics* 2008, **24**(3):133–141.
30. Jeon SA, Park JL, Park SJ, Kim JH, Kim SY: **Comparison between MGI and Illumina sequencing platforms for whole genome sequencing.** *Genes & Genomics* 2021.
31. Jeon SA, Park JL, Kim J-H, Kim JH, Kim YS, Kim JC, Kim S-Y: **Comparison of the MGISEQ-2000 and Illumina HiSeq 4000 sequencing platforms for RNA sequencing.** *Genomics & informatics* 2019, **17**(3).
32. Raine A, Liljedahl U, Nordlund J: **Data quality of whole genome bisulfite sequencing on Illumina platforms.** *PloS one* 2018, **13**(4):e0195972.
33. Wainberg M, Sinnott-Armstrong N, Mancuso N, Barbeira AN, Knowles DA, Golan D, Ermel R, Ruusalepp A, Quertermous T, Hao K: **Opportunities and challenges for transcriptome-wide association studies.** *Nature genetics* 2019, **51**(4):592–599.
34. Deveson IW, Hardwick SA, Mercer TR, Mattick JS: **The dimensions, dynamics, and relevance of the mammalian noncoding transcriptome.** *Trends in Genetics* 2017, **33**(7):464–478.

## Tables

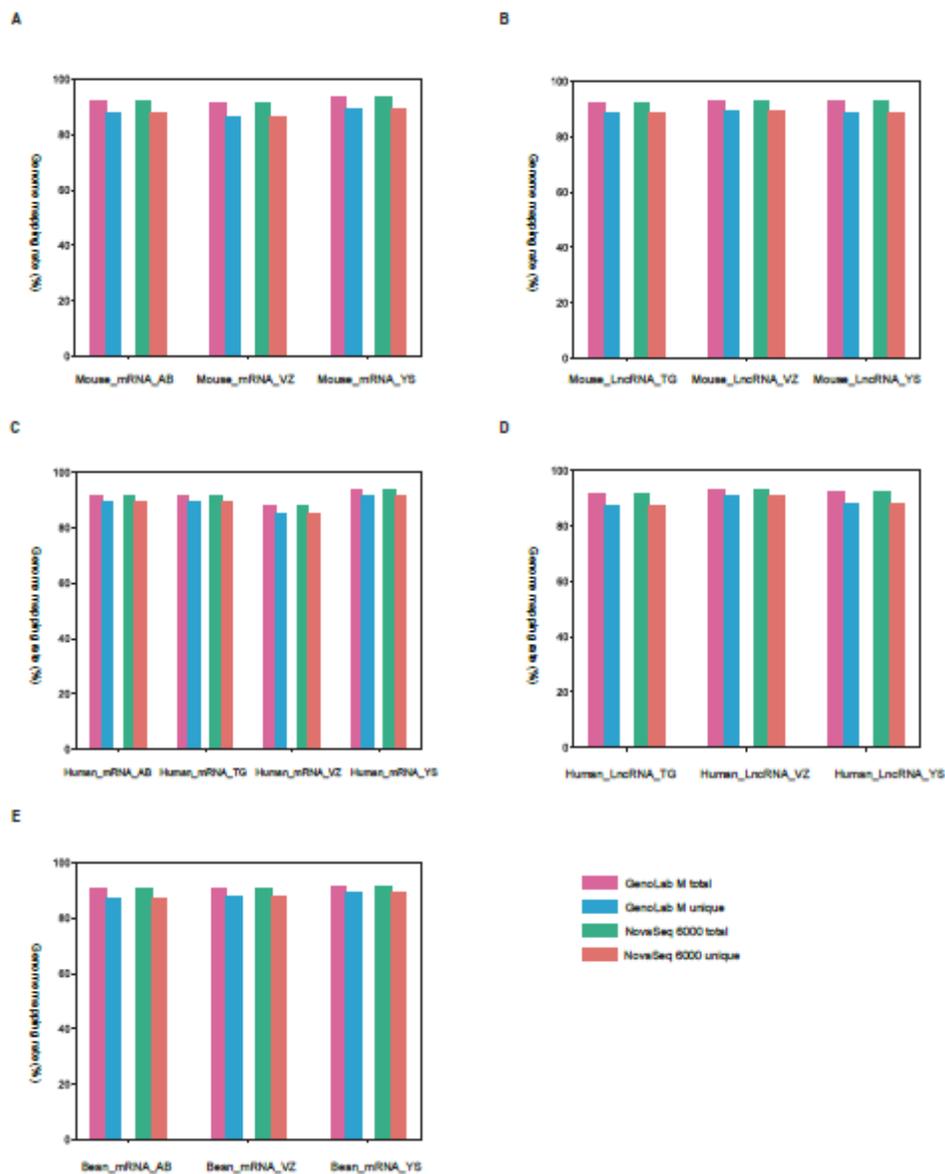
Due to technical limitations, table 1 to 3 is only available as a download in the Supplemental Files section.

## Figures



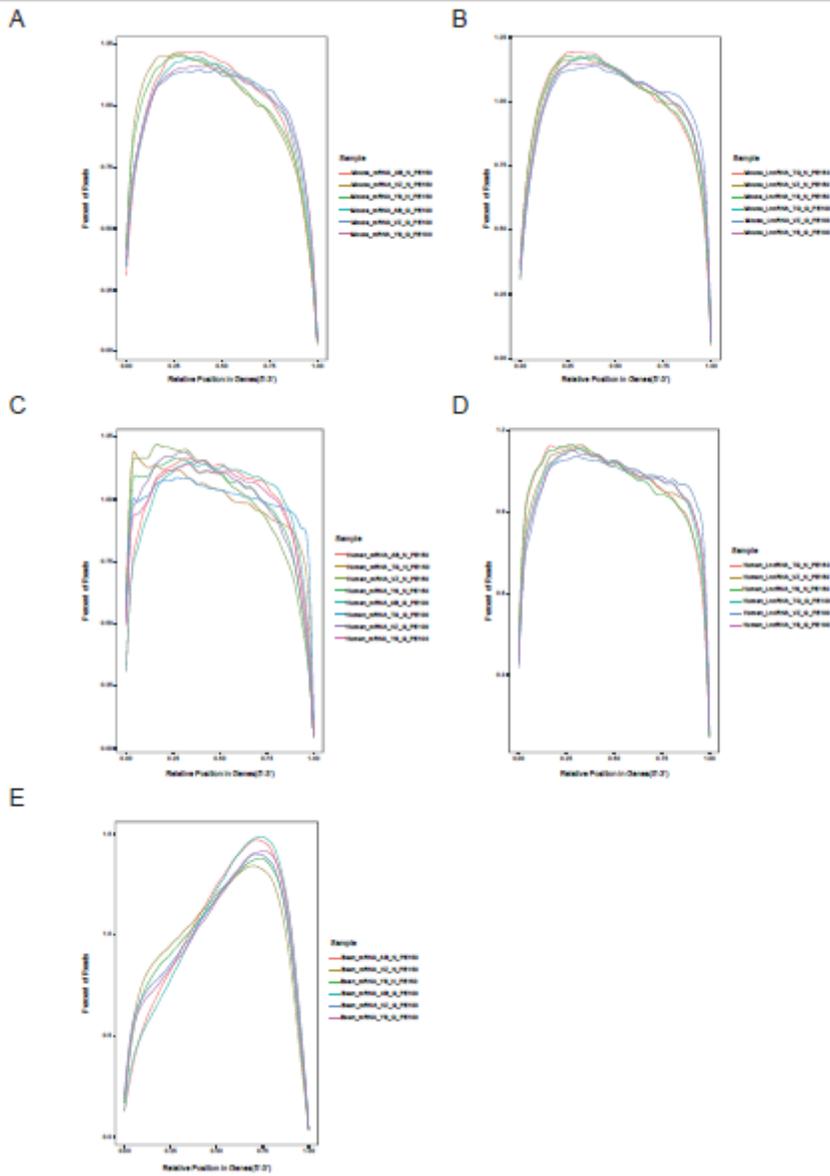
**Figure 1**

Sequencing Workflow of GenoLab M



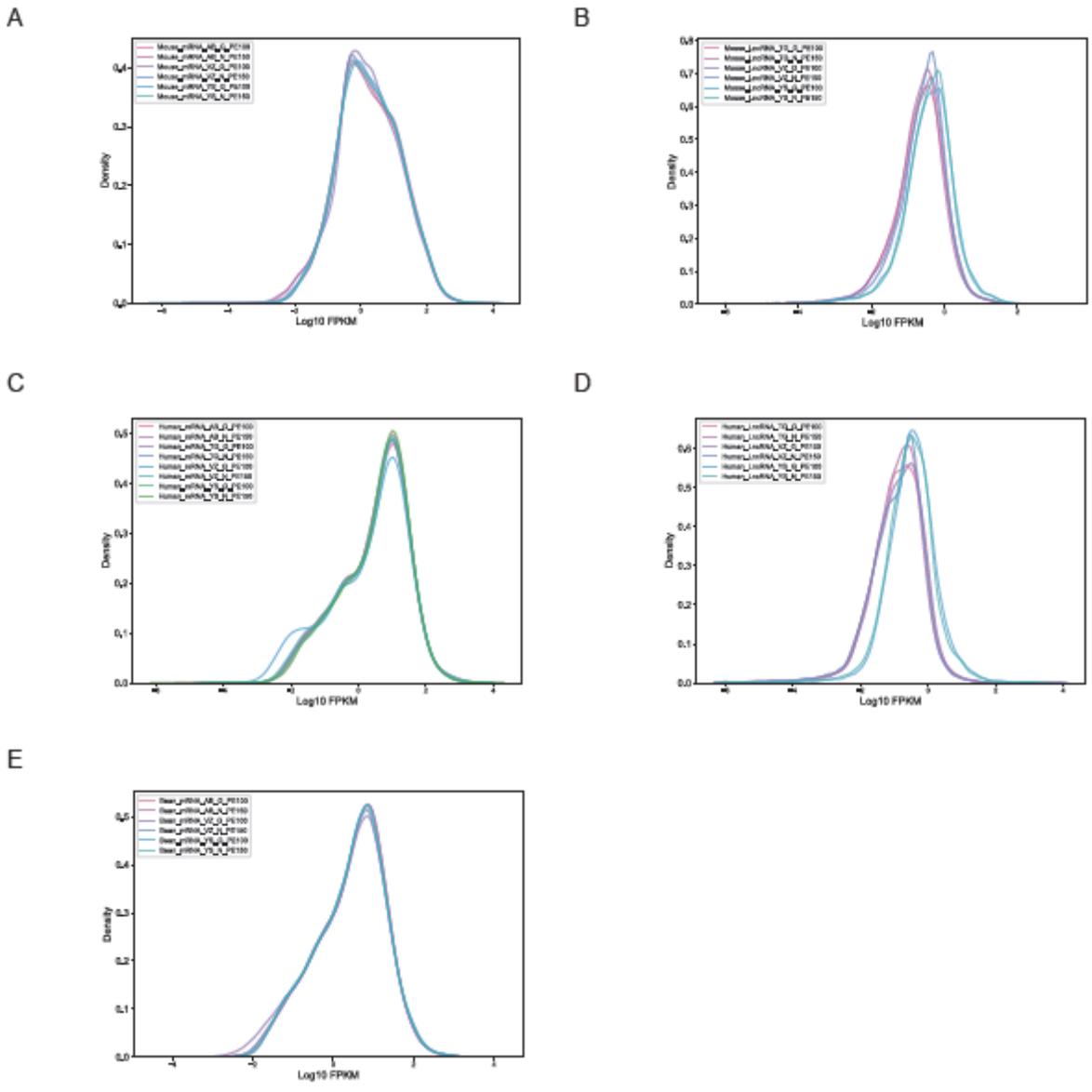
**Figure 2**

Comparison of sequencing quality between GenoLab M and NovaSeq 6000 in genome mapping rate. A Transcriptome of mouse, B LncRNA of mouse, C Transcriptome of human, D LncRNA of human, E Transcriptome of bean.



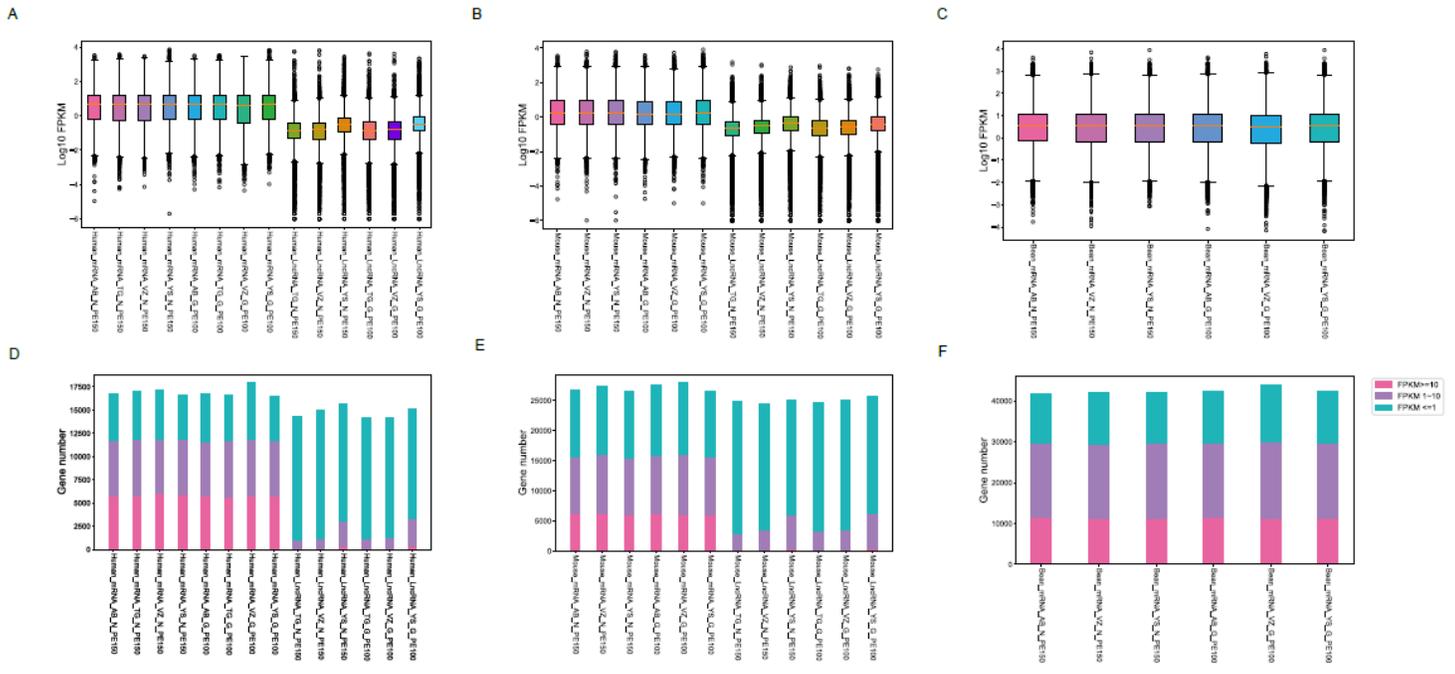
**Figure 3**

Comparison of sequencing quality between GenoLab M and NovaSeq 6000 in Reads distribution along the relative position of genes. A Transcriptome of mouse, B LncRNA of mouse, C Transcriptome of human, D LncRNA of human, E Transcriptome of bean.



**Figure 4**

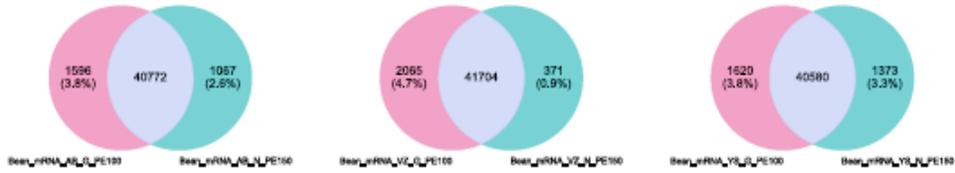
Repeatability of gene detection and quantification between GenoLab M and NovaSeq 6000 in expression density distribution. A Transcriptome of mouse, B LncRNA of mouse, C Transcriptome of human, D LncRNA of human, E Transcriptome of bean.



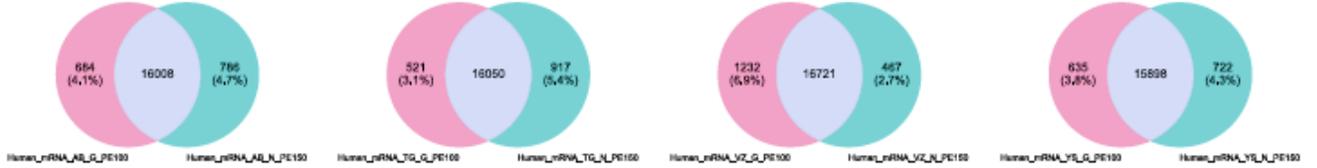
**Figure 5**

Gene and lncRNA detection and quantification between GenoLab M and NovaSeq 6000 in boxplot graph and high and low abundance bar graph. boxplot graph of A human, B mouse, C bean and bar graph of D human, E mouse, F bean.

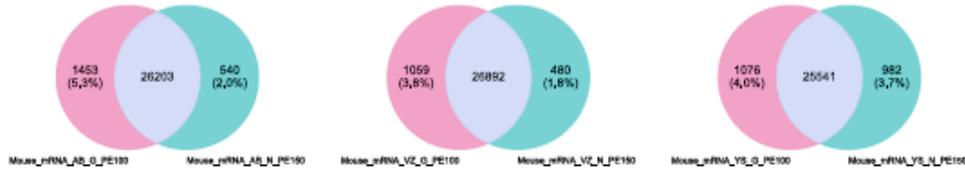
A



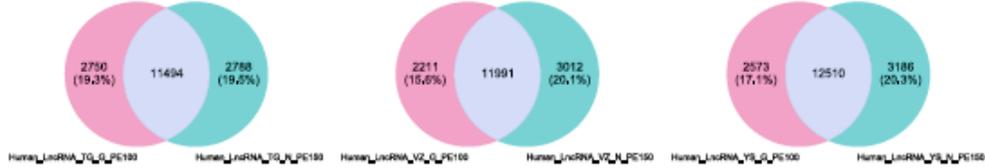
B



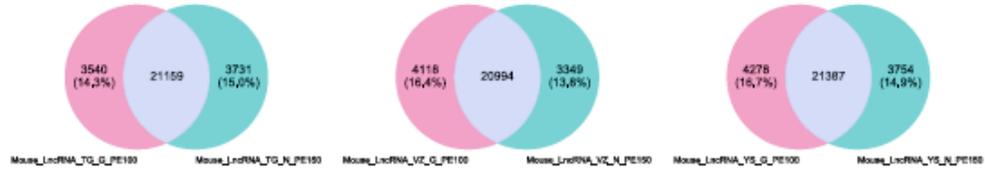
C



D

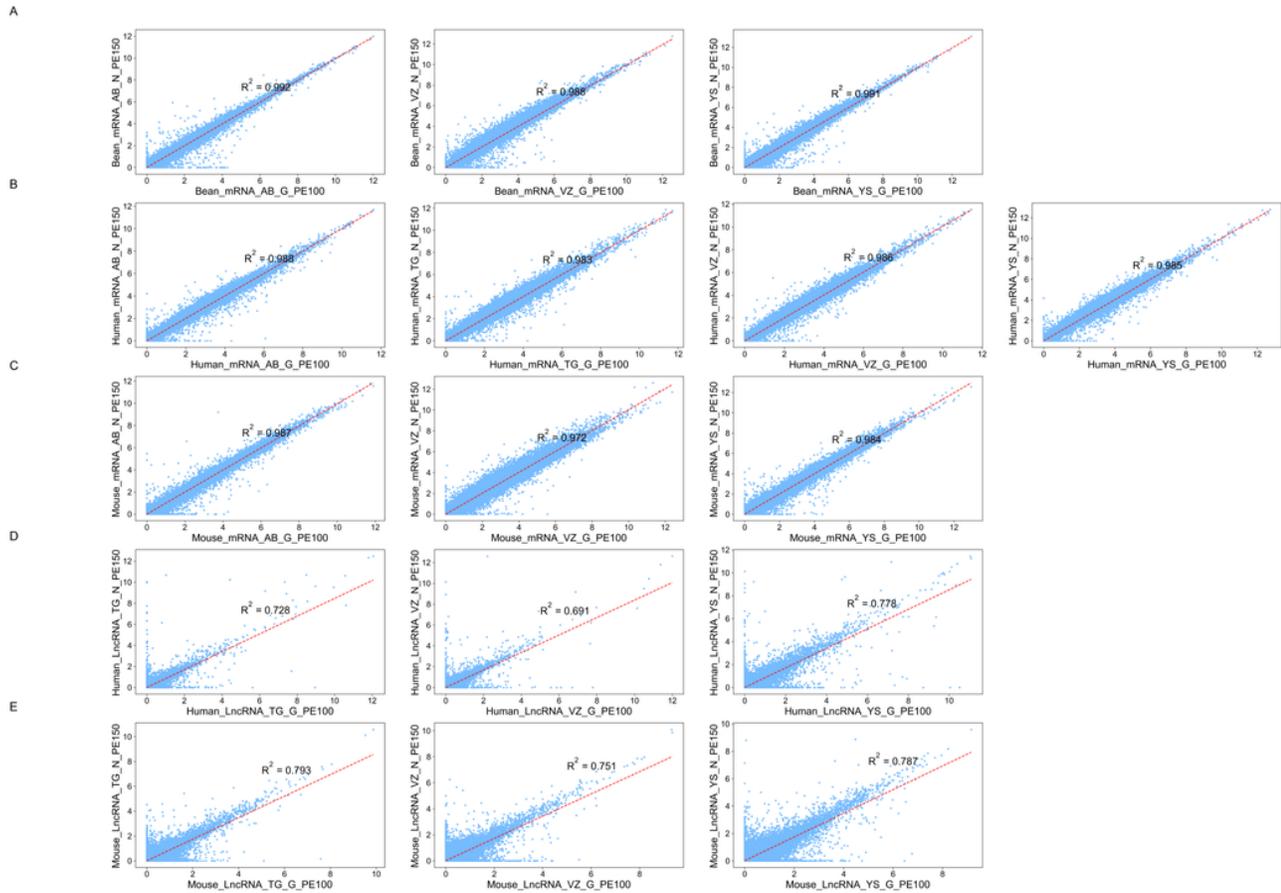


E



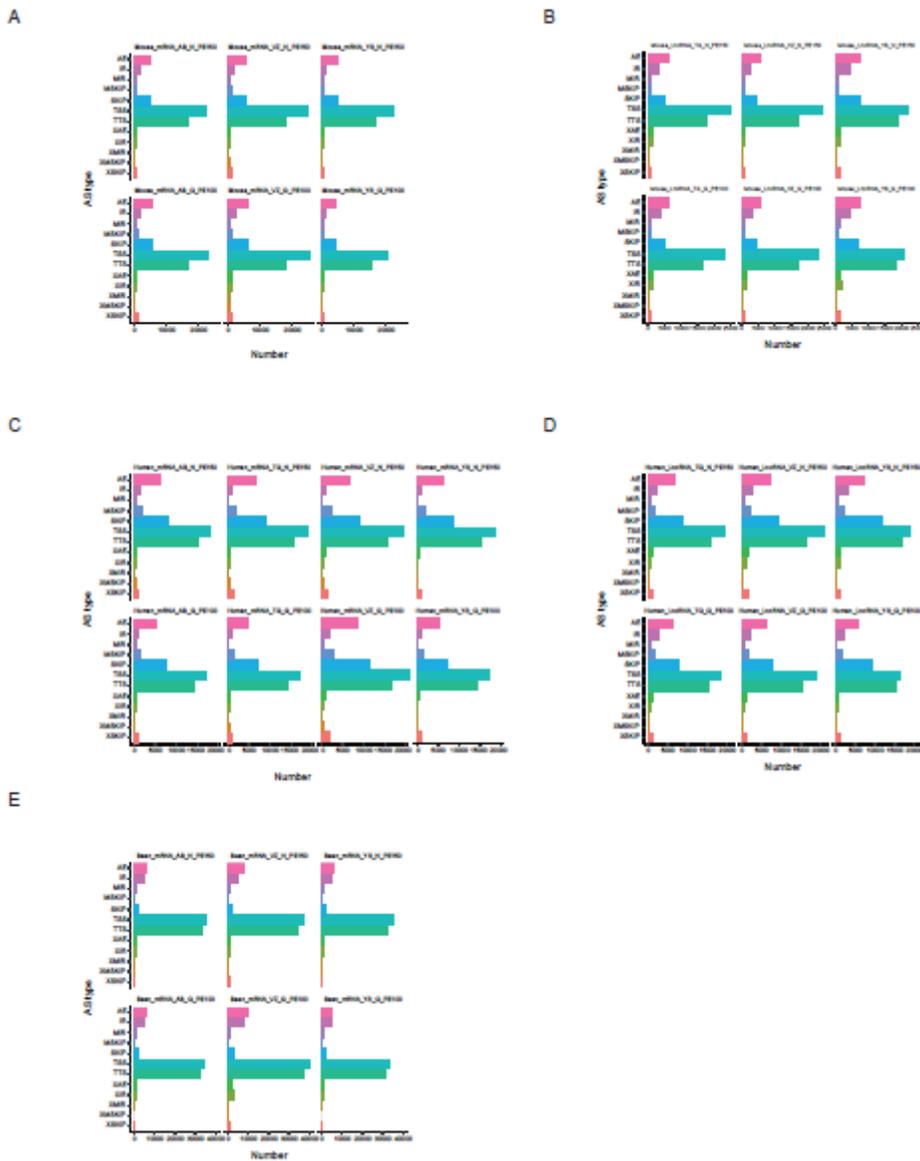
**Figure 6**

Venn diagram of mRNA and lncRNA FPKM between GenoLab M and NovaSeq 6000. A transcriptome of bean, B Transcriptome of human, C Transcriptome of mouse, D lncRNA of human, E lncRNA of mouse.



**Figure 7**

Scatter plots of gene expression values of the four pairs of samples produced using the NovaSeq 6000 and GenoLab M sequencers. Gene expression values are represented as the base 2 logarithm of FPKM. The Pearson correlation coefficients of the 16 samples were between 0.69 and 0.99. A transcriptome of bean, B Transcriptome of human, C Transcriptome of mouse, D lncRNA of human, E lncRNA of mouse.



**Figure 8**

Alternative splicing events of mRNA and lncRNA FPKM between GenoLab M and NovaSeq 6000. A transcriptome of bean, B Transcriptome of human, C Transcriptome of mouse, D LncRNA of human, E LncRNA of mouse. TSS: Alternative 5' first exon (transcription start site), TTS: Alternative 3' last exon (transcription terminal site), SKIP: Skipped exon(SKIP\_ON,SKIP\_OFF pair) , XSKIP: Approximate SKIP (XSKIP\_ON,XSKIP\_OFF pair), MSKIP: Multi-exon SKIP (MSKIP\_ON,MSKIP\_OFF pair), XMSKIP: Approximate MSKIP (XMSKIP\_ON,XMSKIP\_OFF pair), IR: Intron retention (IR\_ON, IR\_OFF pair), XIR: Approximate IR (XIR\_ON,XIR\_OFF pair), MIR: Multi-IR (MIR\_ON, MIR\_OFF pair), XMIR: Approximate MIR (XMIR\_ON, XMIR\_OFF pair), AE: Alternative exon ends (5', 3', or both), XAE: Approximate AE.

## Supplementary Files

This is a list of supplementary files associated with this preprint. [Click to download.](#)

- [Table1.xlsx](#)
- [Table2.xlsx](#)
- [Table3.xlsx](#)
- [SupplementalTableS1.xlsx](#)