SQANTI3 report
Unique Genes: 24102
Unique Isoforms: 66538

<table>
<thead>
<tr>
<th>Category</th>
<th>Genes, count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annotated Genes</td>
<td>14347</td>
</tr>
<tr>
<td>Novel Genes</td>
<td>9755</td>
</tr>
</tbody>
</table>

Gene Classification

<table>
<thead>
<tr>
<th>Category</th>
<th>Isoforms, count</th>
</tr>
</thead>
<tbody>
<tr>
<td>FSM</td>
<td>10525</td>
</tr>
<tr>
<td>ISM</td>
<td>9576</td>
</tr>
<tr>
<td>NIC</td>
<td>3335</td>
</tr>
<tr>
<td>NNC</td>
<td>26170</td>
</tr>
<tr>
<td>Genic Genomic</td>
<td>2599</td>
</tr>
<tr>
<td>Antisense</td>
<td>713</td>
</tr>
<tr>
<td>Fusion</td>
<td>4244</td>
</tr>
<tr>
<td>Intergenic</td>
<td>9376</td>
</tr>
<tr>
<td>Genic Intron</td>
<td>0</td>
</tr>
</tbody>
</table>
## Splice Junction Classification

<table>
<thead>
<tr>
<th>Category</th>
<th>SJs, count</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Known canonical</td>
<td>134052</td>
<td>70.82%</td>
</tr>
<tr>
<td>Known Non–canonical</td>
<td>110</td>
<td>0.06%</td>
</tr>
<tr>
<td>Novel canonical</td>
<td>49152</td>
<td>25.97%</td>
</tr>
<tr>
<td>Novel Non–canonical</td>
<td>5978</td>
<td>3.16%</td>
</tr>
</tbody>
</table>
Gene Characterization
Number of Isoforms per Gene

Genes, %

Isoforms per gene

1
2-3
4-5
>=6

Genes, %
Number of Isoforms per Gene

Known vs Novel Genes

Annotated Genes

Novel Genes

Isoforms Per Gene

- 1
- 2-3
- 4-5
- >=6
Distribution of Mono− vs Multi−Exon Transcripts

Transcripts, %

Novel Genes

Annotated Genes

Transcript type

- Multi−Exon
- Mono−Exon
**Structural Categories by Transcript Length**

Transcript length, kb

Structural Category
- FSM
- ISM
- NIC
- NNC
- Genic
- Genomic
- Antisense
- Fusion
- Intergenic
- Intron

%
Annotated vs Novel Gene Expression

log2(Gene_TPM+1)

Annotated Genes

Novel Genes
Transcript Lengths Distribution by Structural Category

- FSM
- NIC
- Genic
- Genomic
- Fusion
- Genic Intron
- ISM
- NNC
- Antisense
- Intergenic

Count

Transcript length

0 4000 8000 12000
Mono- vs Multi- Exon Transcript Lengths Distribution

Transcript length

Count

exonCat

Multi-Exon
Mono-Exon
Structural Isoform Characterization
Intron retention

At least 1 annot. don./accept.

Isoform Distribution Across NNC

Coding prediction
- Coding
- Non coding

Transcripts, %
Isoform Distribution Across NIC

- Comb. of annot. splice sites
- Intron retention

Coding prediction:
- Coding
- Non coding
Transcript Lengths by Subcategory

- Alternative 3'end
- Alternative 3'5'end
- Alternative 5'end
- Reference match
- 3' fragment
- Internal fragment
- 5' fragment
- Comb. of annot. junctions
- Comb. of annot. splice sites
- Intron retention
- Not comb. of annot. junctions
- Mono-exon by intron ret.
- At least 1 annot. don./accept.
- Mono-exon
- Multi-exon
Exon Counts by Structural Classification

- FSM
- ISM
- NIC
- NNC
- Genic
- Genomic
- Antisense
- Fusion
- Intergenic
- Genic Intron
Transcript Expression by Subcategory

- Genic
- Genomic
- Antisense
- Fusion
- Intergenic

- log2(TPM+1)

- Alternative 3'end
- Alternative 3'5'end
- Alternative 5'end
- Reference match
- Internal fragment
- 3' fragment
- Internal fragment
- 5' fragment
- Comb. of annot. junctions
- Comb. of annot. splice sites
- Intron retention
- Not comb. of annot. junctions
- Mono–exon by intron ret.
- At least 1 annot. don./accept.
- Mono–exon
- Multi–exon
Length Distribution of Matched Reference Transcripts

Applicable Only to FSM and ISM Categories
Exon Count Distribution of Matched Reference Transcripts

Applicable Only to FSM and ISM Categories
Gene Expression of NNC And Not NNC Containing Genes

log2(Gene TPM+1)
Splice Junction Characterization
Distribution of Splice Junctions by Structural Classification

Splice junctions, %


Splice junctions by structural classification.
Distribution of Transcripts by Splice Junctions

Transcripts, %

FSM | ISM | NIC | NNC | Genic Genomic | Antisense | Fusion | Intergenic

Canonical | Non–canonical

Graph showing the distribution of transcripts by splice junctions, with categories including FSM, ISM, NIC, NNC, Genic Genomic, Antisense, Fusion, and Intergenic.
RT-Switching All Junctions

- Known canonical: 0.34%
- Known non-canonical: 0%
- Novel canonical: 0.46%
- Novel non-canonical: 14.34%
Unique Junctions RT–switching

RT–switching junctions, %

- **Known canonical**: 1.53%
- **Known Non–canonical**: 0%
- **Novel canonical**: 1.8%
- **Novel Non–canonical**: 19.09%
Comparison With Annotated TSS and TTS
Distance to annotated Transcription Termination Site (TTS)
FSM

Negative values indicate upstream of annotated termination site
Distance to annotated Transcription Termination Site (TTS) FSM

Negative values indicate upstream of annotated termination site
Distance to Annotated Transcription Start Site for FSM

Negative values indicate downstream of annotated TSS
Distance to Annotated Transcription Start Site for FSM

Negative values indicate downstream of annotated TSS

Distance to annotated Transcription Start Site (TSS), bp
Distance to Annotated Polyadenylation Site for ISM

Negative values indicate upstream of annotated polyA site

polyA motif found

FALSE
Distance to Annotated Polyadenylation Site for ISM

Negative values indicate upstream of annotated polyA site
Distance to Annotated Transcription Start Site for ISM

Negative values indicate downstream of annotated TSS
Distance to Annotated Transcription Start Site for ISM

Negative values indicate downstream of annotated TSS

Distance to annotated transcription start site, bp
Comparison With Annotated TSS and TTS by Subcategories
PolyA Distance Analysis
Redundancy Analysis
Reference Transcript Redundancy

Count of Reference Transcripts

Total ISM per reference ID

Only ISM

1784

1675

Unique

Multiple

ISM per reference transcript

1

2

3

4

5

6

7

8+

0

500

1000

1500

2000
Reference Transcript Redundancy

FSM+ISM per reference transcript

- Unique: 3786
- Multiple: 3967

Total FSM+ISM per reference transcript:
- 1: 1
- 2: 1
- 3: 1
- 4: 1
- 5: 1
- 6: 1
- 7: 1
- 8+: 1
Intra–Priming Quality Check
Possible Intra–Priming by Structural Category

Percent of genomic 'A's in downstream 20 bp

- Alternative 3'end
- Alternative 3'5'end
- Alternative 5'end
- Reference match
- 3' fragment
- Internal fragment
- 5' fragment
- Comb. of annot. junctions
- Comb. of annot. splice sites
- Intron retention
- Not comb. of annot. junctions
- Mono–exon by intron ret.
- At least 1 annot. don./accept.
- Mono–exon
- Multi–exon
Possible Intra-Priming by Structural Category

Percent of genomic 'A's in downstream 20 bp
Possible Intra-Priming by Structural Category

Percent of genomic 'A's in downstream 20 bp
Mono- vs Multi-Exon Possible Intra-Primming

Percent of genomic 'A's in downstream 20 bp

'Multi Exon Isoforms' vs 'Mono Exon Isoforms'

- FSM
- ISM
- NIC
- NNC
- Genic Genomic
- Antisense
- Fusion
- Intergenic

'Multi Exon Isoforms' show higher percent of genomic 'A's in downstream 20 bp compared to 'Mono Exon Isoforms' in most categories.
Coding vs Non–Coding Possible Intra–Priming

Percent of genomic 'A's in downstream 20 bp

'C'A'S, %

FSM  ISM  NIC  NNC  Genic  Genomic  Antisense  Fusion  Intergenic

Coding Isoforms  Non–Coding Isoforms
Features of Bad Quality
RT–switching

Isoforms, %

- FSM: 5.7%
- NIC: 2.9%
- NNC: 7.2%
Non-Canonical Junctions

17.5%
Splice Junctions Without Short Read Coverage

- FSM: 0.5%
- ISM: 0.8%
- NIC: 1%
- NNC: 20.4%
Nonsense–Mediated Decay by Structural Category

- FSM: 0.2%
- ISM: 2.3%
- NIC: 30.2%
- NNC: 14.2%
Quality Control Attributes Across Structural Categories

Transcripts, %

Non-canonical
Not Coverage SJ
Predicted NMD
RT switching
Features of Good Quality
Splice Junctions With Short Read Coverage

- FSM: 99.5%
- ISM: 99.2%
- NIC: 99%
- NNC: 79.6%
Good Quality Control Attributes Across Structural Categories

Transcripts, %

FSM
ISM
NIC
NNC

Annotated
Canonical
Coverage SJ