detection of rearrangements in cancer genome with ChIP-seq data
structural rearrangements

types:

- deletion: A B E F G
- inversion: A B C E D F G
- insertion: A B C T U D E F G
- duplication: A B C D E F E F G
- translocation: A B C D W X Y

inspired by http://images.slideplayer.com/15/4829724/slides/slide_2.jpg; acess: 24.01.2018
motivation

rearrangement - cancer causes:
- mutation of tumor suppressor gene
- activation of oncogene
- metastasis and drug resistance¹

detection of rearrangements:
- DNA-PET (Paired-end tag)
  - need material
  - costs time and money²
- this method
  - only use already determined data
  - freely available data

basic idea

ChIP-seq read → mapping → reference genome

breakpoints

split read
basic idea

ChIP-seq read

mapping

reference genome

5'

“normal” read

3'
first part of the workflow - preparation

- Provide data
  - from Encode
  - control data
  - K562 cancer
  - 27,494,399 read

- Quality check
  - with fastqc
  - if necessary cut adapter

- Mapping
  - with segemehl
  - allow split reads
  - hg38 reference genome
extraction of breakpoints

stop values: ...0 0 0 0 0 1 0 3 0 0...

“normal” reads
split reads
extraction of breakpoints

normal reads

split reads

breakpoint interval (merged)

pass values: ...1 2 3 5 5 4 4 1 1 1...

stop values: ...0 0 0 0 1 0 3 0 0...
extraction of breakpoints

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

pass values: ...
stop values: ...

add up pass values within

<p>| |</p>
<table>
<thead>
<tr>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>100 pos. + interval</td>
</tr>
<tr>
<td>200 pos. + interval</td>
</tr>
<tr>
<td>400 pos. + interval</td>
</tr>
<tr>
<td>whole genome</td>
</tr>
</tbody>
</table>
calculation of the p-values

[Diagram showing a bar graph with x-axis labeled 'value' and y-axis labeled 'probability'. The values range from 2 to 10 on the x-axis, and probability values range from 0 to 0.20. The bar for the stop-value is marked and the bars for p-values decrease as the value increases.]
following tasks

- p-value limit
- simple or complex

- accuracy
- practicability
- improvements
- disadvantages
thank you for your attention