

# Visualisation for *segemehl* output

Hermann Pauly

- *segemehl*: alignment tool for NGS
- can map strand switching and circular events
- no software to plot output yet
- tool to create plots from *segemehl* mappings
- allow search for "unusual" splice events
- user-defined selection and filtering of plots
- graph theory
- on-screen and file output
- C++, Qt4, .eps
- test data
  - simulated chromosomes
  - manually created RNA-reads
  - re-mapped with *segemehl*

- real data
  - RNA sequences of human skin cells
  - already mapped by *segemehl*
- exons as nodes
- splicings as edges
- chromosomes: maps of links to nodes
- chromosomes (name, length, exon pos)
- exons for plots (id, length, position)
- connection information



data, 3' -> 5')

(test

	index	% time	self	children	called	name
Visualisation for <i>segemehl</i> output						
Hermann Pauly	[1]	97.6	0.00	3.27		main
	[2]	96.9	0.00	3.24	1	Ge
Introduction	[3]	96.9	0.01	3.23	1	Ge
Goal	[4]	84.2	0.08	2.74	199914	Ge
My approach						
Read as graph						
Datastructure: reassembly			0.00	0.00	253/200167	
Datastructure: plotting			0.05	1.66	199914/200167	
Results	[5]	51.1	0.05	1.66	200167	st
Multisplit plot			0.06	1.04	3579101/3579185	
Profiling			0.00	0.24	200167/200167	
Problems and outlook			0.03	0.09	200167/200167	
Problems			0.04	0.07	3779268/3779268	
To do / planned			0.03	0.04	3579101/3579101	
Sources			0.00	0.03	200167/200167	
strsplit()						

## Visualisation for *segemehl* output

Hermann  
Pauly

### Introduction Goal

### My approach

Read as graph  
Datastructure:  
reassembly  
Datastructure:  
plotting

### Results

Multisplit plot  
Profiling

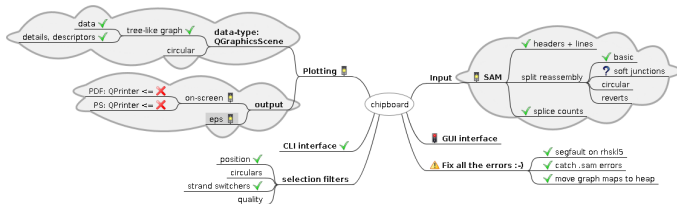
### Problems and outlook

Problems  
To do / planned

### Sources

strsplit()

0.00	0.00	200167/200167
0.00	0.00	200167/200168



- *segemehl* doesn't honour *.sam* standard
- memory: graph of 40+ GiB input files
- data not ordered
- Qt4 file printing surprisingly bad
- visuals
  - include strandiness information

- shortened chromosome displays
- more information
- function
  - more filtering
  - circular elements
  - graphical interface
  - more input formats
  - include genomic annotation

Thank you

```
vector<string> strsplit ( string& input, string  
    string token, theStr(input);  
    int L = delim.length();  
    vector<string> result(new vector<string>());  
  
    while (token != theStr) {  
        auto end = theStr.find_first_of(delim);
```

```
token = theStr.substr(0, end);  
theStr = theStr.substr(end + L);  
if (keepEmpty || token.length() > 0) {  
    result.push_back(token);  
}  
}  
return result;  
}
```