

Detection of viral integration sites

Corinna Blasse

Advanced Algorithms for Bioinformatics (P4)
Instructor: Birte Kehr

SoSe 2010
FU Berlin

July 18, 2010

Biological situation

- ▶ **Human Herpesvirus Six** (HHV-6) is a wide spread virus
- ▶ One possible persistence by virus integration



- ▶ Virus integration: Insertion of viral DNA into host-cell DNA

Experiment

- ▶ Integration of human herpesvirus 6 into the human genome
- ▶ Sequencing with the SOLID 3 Plus System



- ▶ **Goal:** To find reads containing virus and host genome



Outline of the project

- ▶ Input:
 - ▶ Millions of SOLID reads
 - ▶ Sequence of the host genome
 - ▶ Sequence of the viral genome
- ▶ Steps of the project:
 - ▶ Develop a local read mapper for color-space reads
 - ▶ Apply the read mapper to find viral integration sites in the human genome

▶ Output:

```
>read14123|chr17:1943124|virus:2535  
AAACCACCACCAGGCGACGC|CGGCAGGGAACGCTTCAGCAT  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
AAACCACCAGCAGGAGACGC|CGGCAGGGAACGCTTCTGCAT
```

Color-space reads

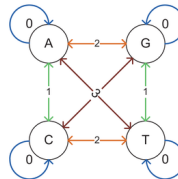
- = Reads that contain one of four colors for every pair of nucleotides

T0001201030

- Encoding by using the AB SOLiD sequencing system

	A	C	G	T
A	0	1	2	3
C	1	0	3	2
G	2	3	0	1
T	3	2	1	0

A



B

Local Read Mapper for color-space reads

1. Import the color-space reads and convert them
2. Import the sequences of the virus and human genome
3. Read mapping by using Swift-Local in SeqAn
4. Filter the mapped reads
 - ▶ Complete mapped reads
 - ▶ Local mapped reads
5. Construct reads that belong to both genomes

Results

- ▶ Mapping of a test data set containing reads of the following structures
 - ▶ Human genome
 - ▶ Virus genome
 - ▶ Virus genome | Human Genome
 - ▶ Human Genome | Virus Genome
- ▶ Just handling of sequence errors, not errors in the color-code

Sources



Principles of Di-Base Sequencing and the Advantages of Color Space Analysis in the SOLiDTM System.

Applied Biosystems 2008.



Rumble SM & Lacroute P & Dalca AV & Fiume M & Sidow A et.al.

SHRiMP: Accurate Mapping of Short Color-space Reads.

PLoS Computational Biology 2009.