

M. Eng. Christian Rohrandt

Prof. Dr.-Ing. Ulrich Jetzek

Faculty of Computer Science and Electrical Engineering

Kiel University of Applied Sciences

**AMIES 2017 Vaasa**

# Recognition of specific short DNA sequence patterns in an electric current stream generated by a nanopore DNA sequencing device

15. September 2017 1/23

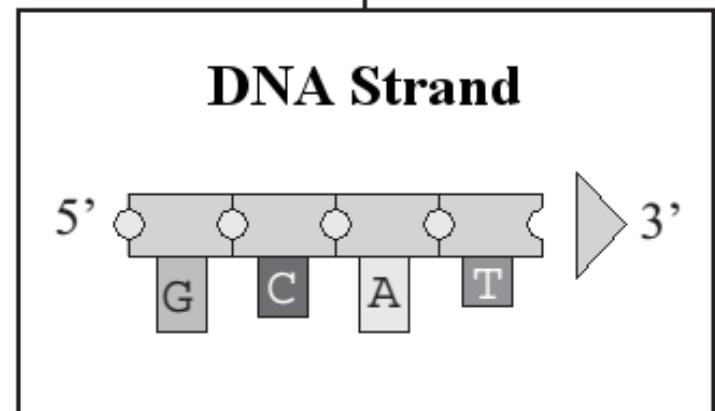
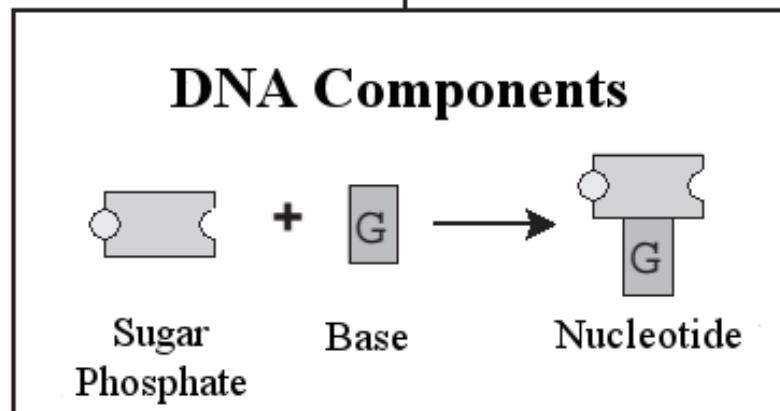
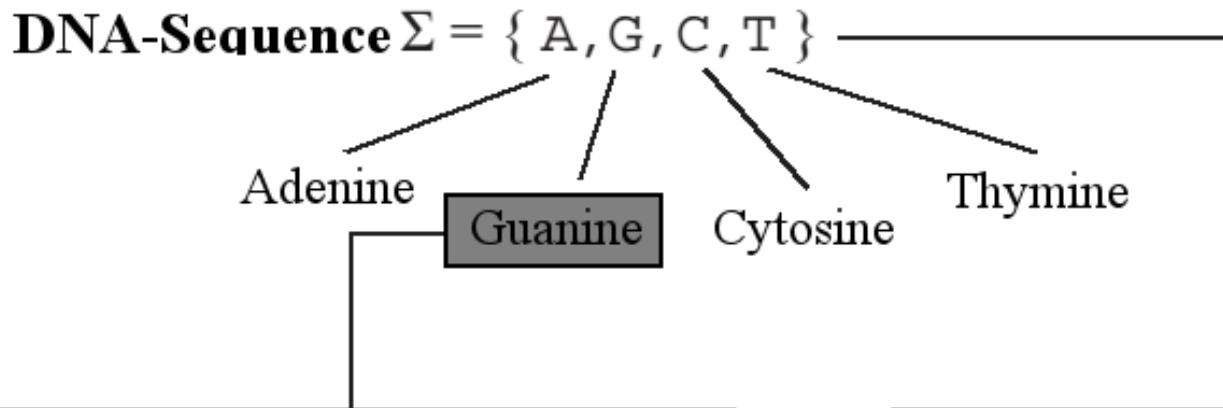


**FACHHOCHSCHULE KIEL**  
University of Applied Sciences

# Content

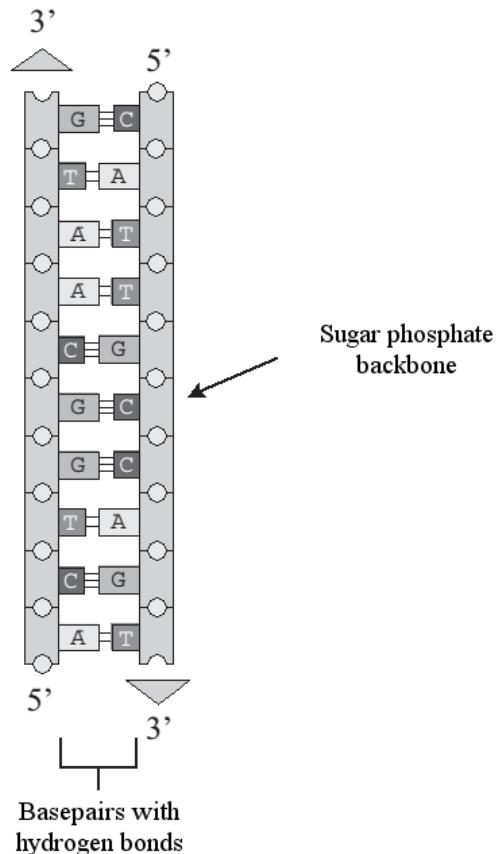
- Introduction
  - What is DNA?
  - How can we read DNA? (Nanopore Sequencing)
  - Oxford Nanopore Technologies MinION Sequencing Device
  - DNA Sequence Alignment
  - Selective DNA Sequencing
- Identifying Patterns in the Current Stream
- Anchoring Patterns in the Reference Genome Sequence
- Intermediate Results
- Future Improvements

# What is DNA? 1/2

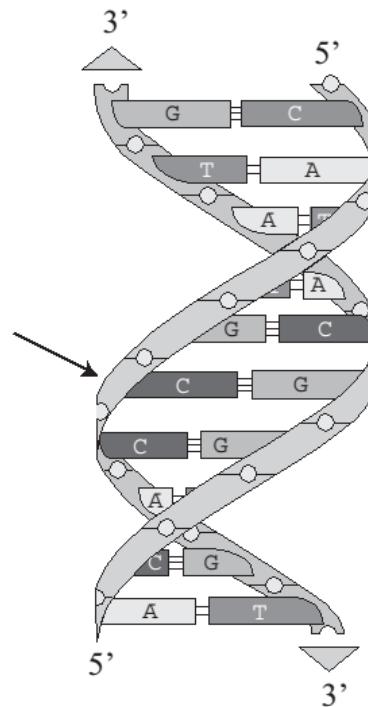


## What is DNA? 2/2

DNA-Doublestrand



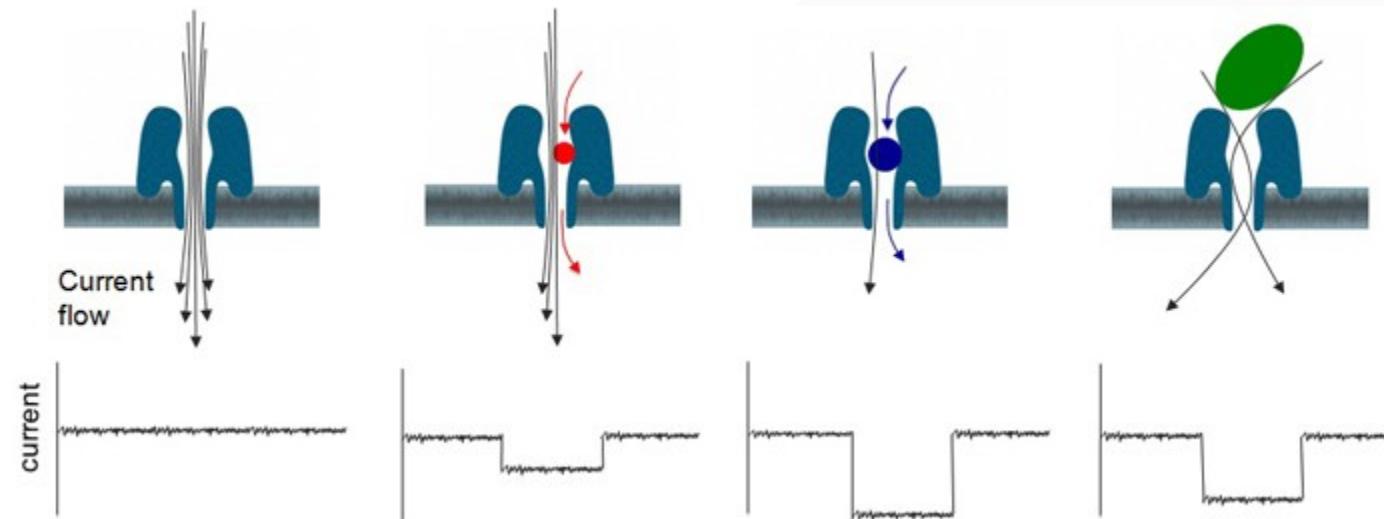
DNA-Doublehelix



M.-T. Hütt and M. Dehnert, Methoden der Bioinformatik. Berlin, Heidelberg:  
Springer Berlin Heidelberg, 2016, Chapter 1.

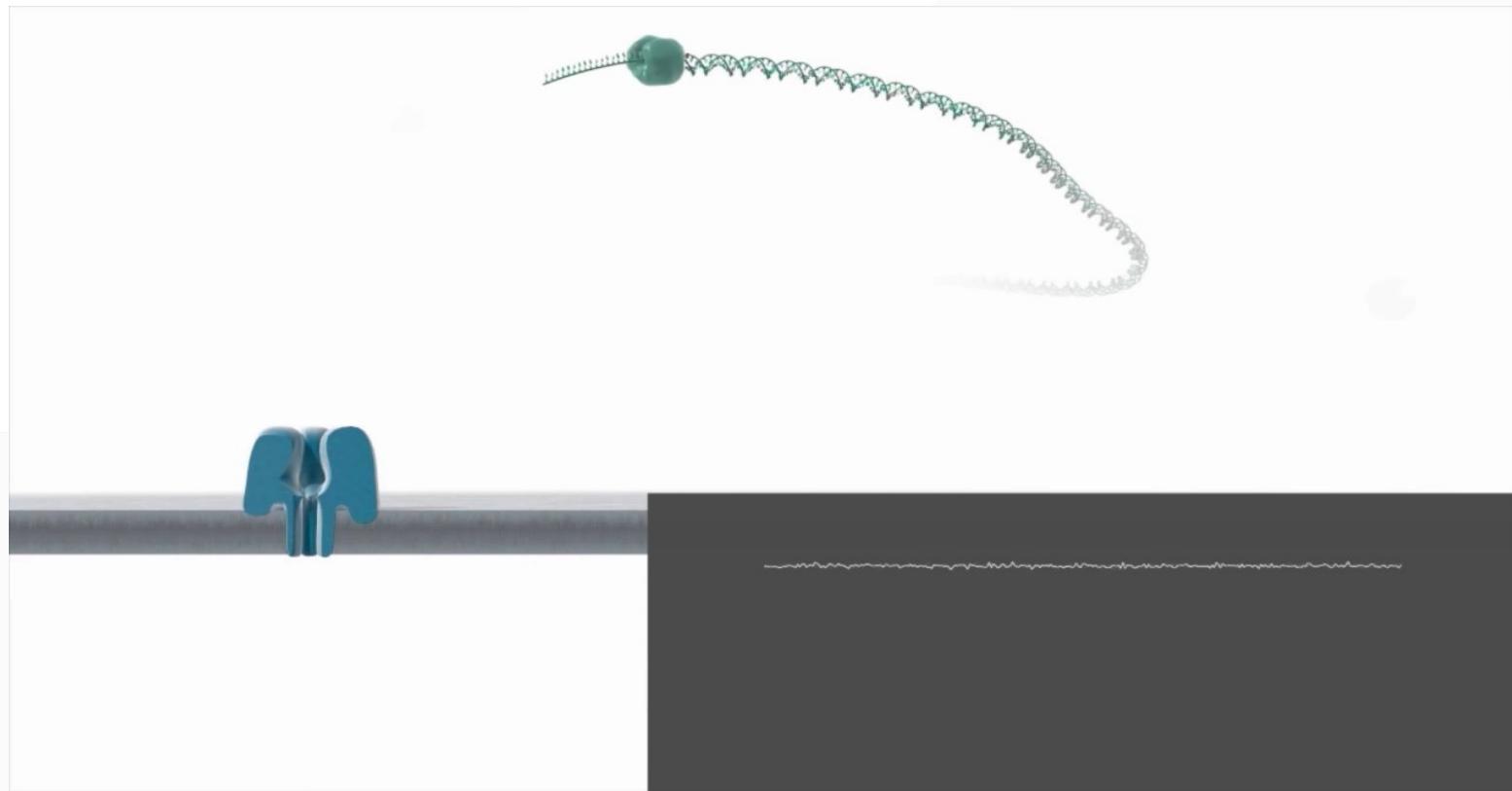
15. September 2017 4/23

# How we can read DNA? (Nanopore Sequencing) 1/2



"Oxford Nanopore unveils portable genome sequencer – MinION." [Online]. Available:  
<https://phys.org/news/2014-02-oxford-nanopore-unveils-portable-genome.html>. [Accessed: 14-Sep-2017] 15. September 2017 5/23

## How we can read DNA? (Nanopore Sequencing) 2/2



Oxford Nanopore Technologies, Nanopore DNA Sequencing. 2015 [Online]. Available:  
<https://www.youtube.com/watch?v=CE4dW64x3Ts>. [Accessed: 13-Sep-2017]

15. September 2017 6/23

# Oxford Nanopore Technologies Sequencing Device 1/2



Oxford Nanopore Technologies, [www.nanoporetech.com](http://www.nanoporetech.com)

15. September 2017 7/23

# Oxford Nanopore Technologies Sequencing Device 2/2

- 2048 nanopores
  - Current measurement using an ASIC in 512 channels
- Sequencing speed
  - max 70 bases/pore/second (R7)
  - max 250 bases/pore/second (R9) (mid 2016)
  - max 450 bases/pore/secind (R9.4) (beginning 2017)
- USB powered
- Portable

# DNA Sequence Alignment



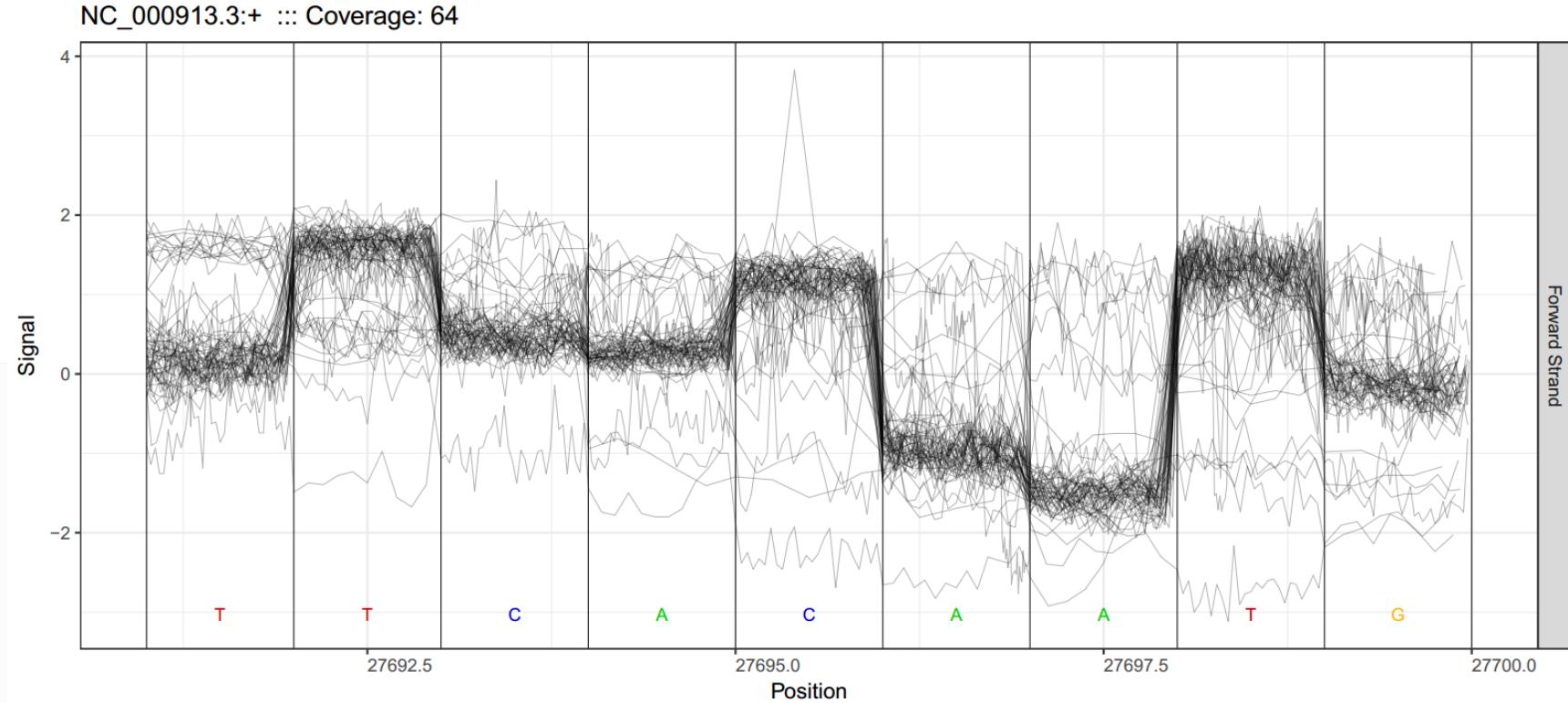
Jaworski, Elizabeth; Routh, Andrew (2017): Comparison of Illumina HiSeq to Oxford Nanopore MinION read data.. figshare. <https://doi.org/10.1371/journal.ppat.1006365.g004> Retrieved: Sep 14, 2017

15. September 2017 9/23

# Selective DNA Sequencing

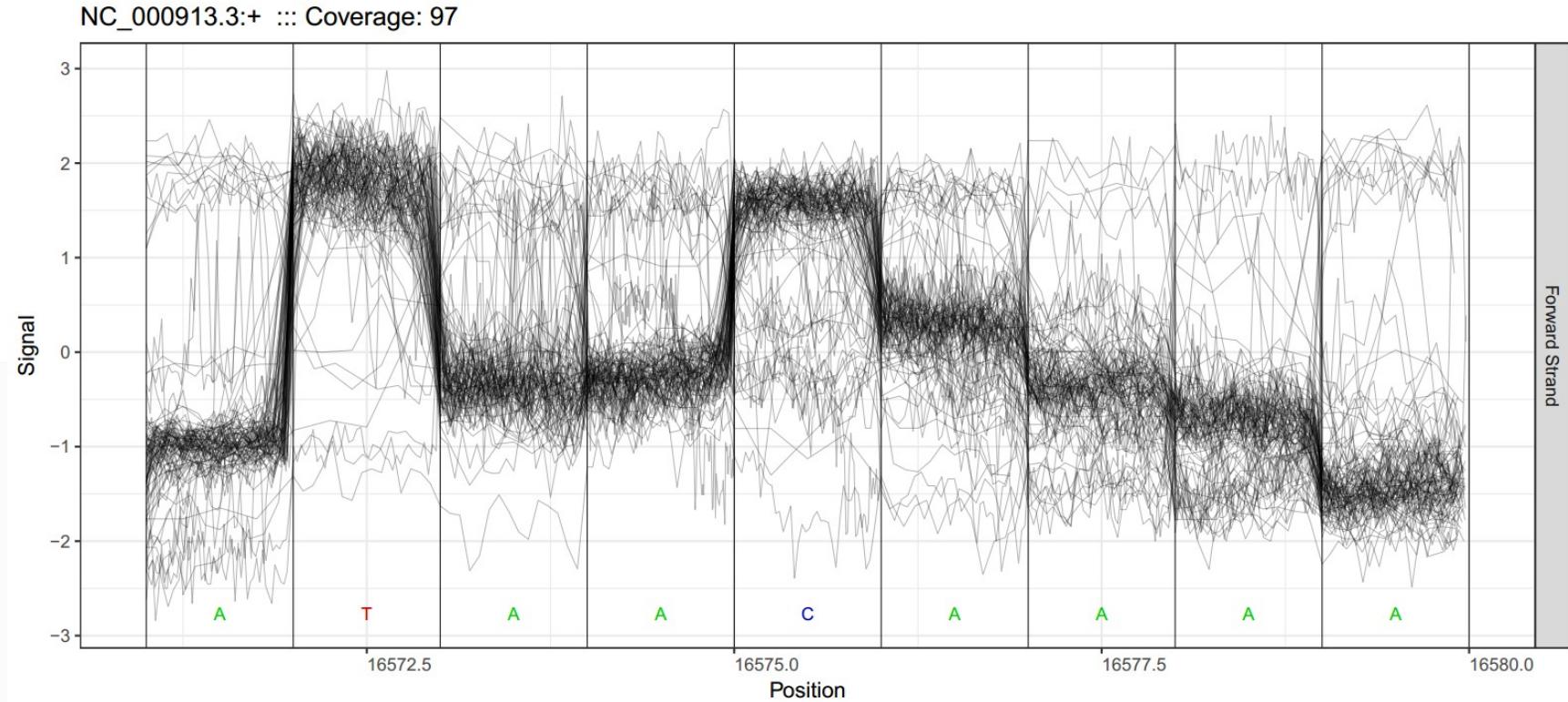
- Problem
  - DNA samples may be contaminated
  - Different species
  - Only small parts of the whole genome of interest
  - Single genes are examined
- Possible solution
  - Locate the current sample within the reference genome
    - or
  - Assign the current sample to a specific species
  - Continue sequencing when positive match
    - or
  - Eject the sample from the pore

# Identifying Patterns in the Current Stream 1/4



15. September 2017 11/23

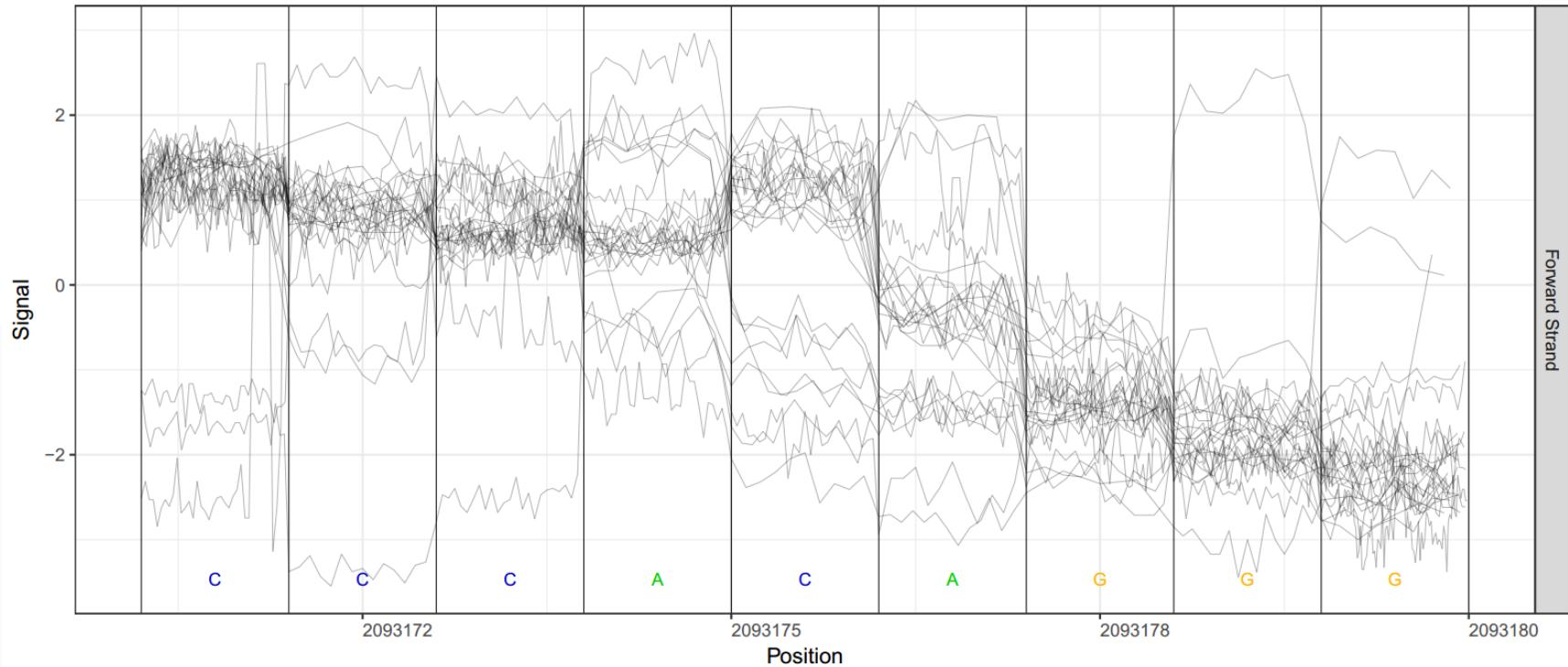
## Identifying Patterns in the Current Stream 2/4



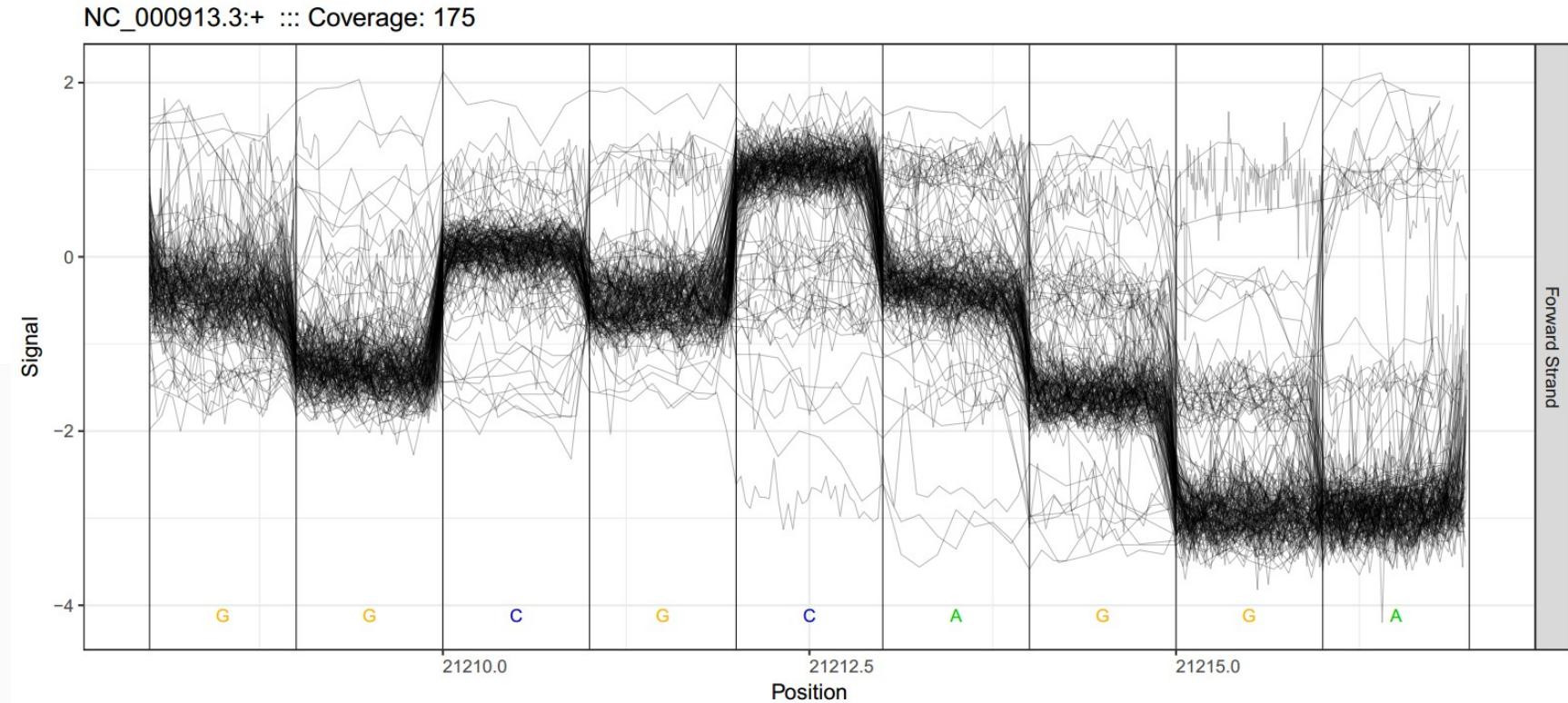
15. September 2017 12/23

## Identifying Patterns in the Current Stream 3/4

NC\_000913.3:+ :: Coverage: 26



# Identifying Patterns in the Current Stream 4/4



15. September 2017 14/23

# Anchoring Patterns in the Reference Genome Sequence

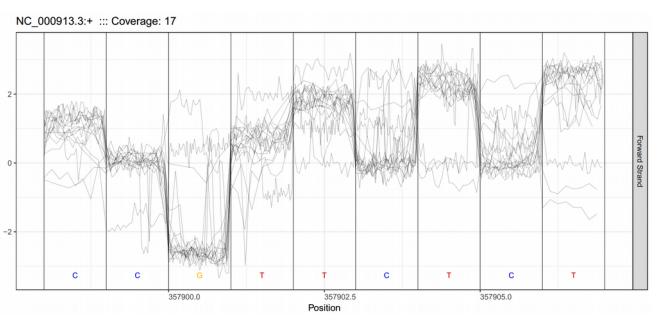
...TTGGCTTGCGTCTGTTCAGCGTGCCTCGCCAGTTCTGTTCAGGCGCAGGAAGTCGTGCTGGACTGGTCACATCAACGTAGATGGATCAAGTTGCTGCACGGTTGCCAGCGCAGTCGCCCTGACC GTTCTGTACCAAT...



15. September 2017 15/23

# Anchoring Patterns in the Reference Genome Sequence

...TTGGCTTG**CCGTTCTCT**TGTTTCAAGCGTGCCATTGCCAGTCCTGTTCAAGCGCAGGAAGTCGTGCTGGACTGGTCACATCAACGTAGATGGATCAAGTTGCTGCACGGTTGCCAGCGCAGTCGCCCTGACCCTGTAACCAAT...

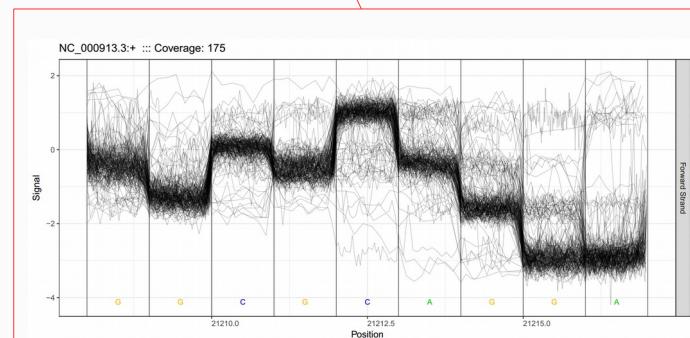
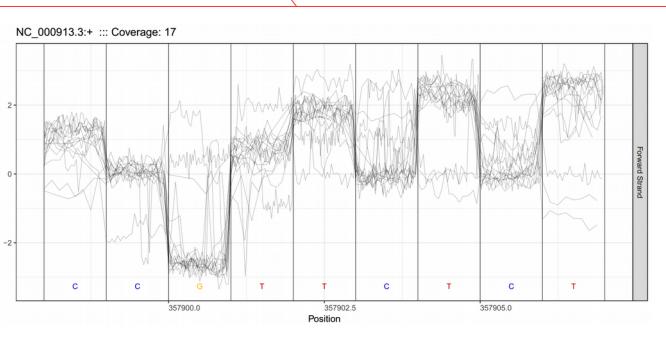


...GCGCCCGCAATAATCACCTGATAAC**CCGTTCTCT**TCGGCGCTTCGGCGAAGCTGAACAGTTATCGGGGGTGCCTGAGCAGAAACCACTAACGTGGTGCAGGACATTCAAGGATTTCGAAGATTTCGGCGCGAAGTCATGGTAGC...

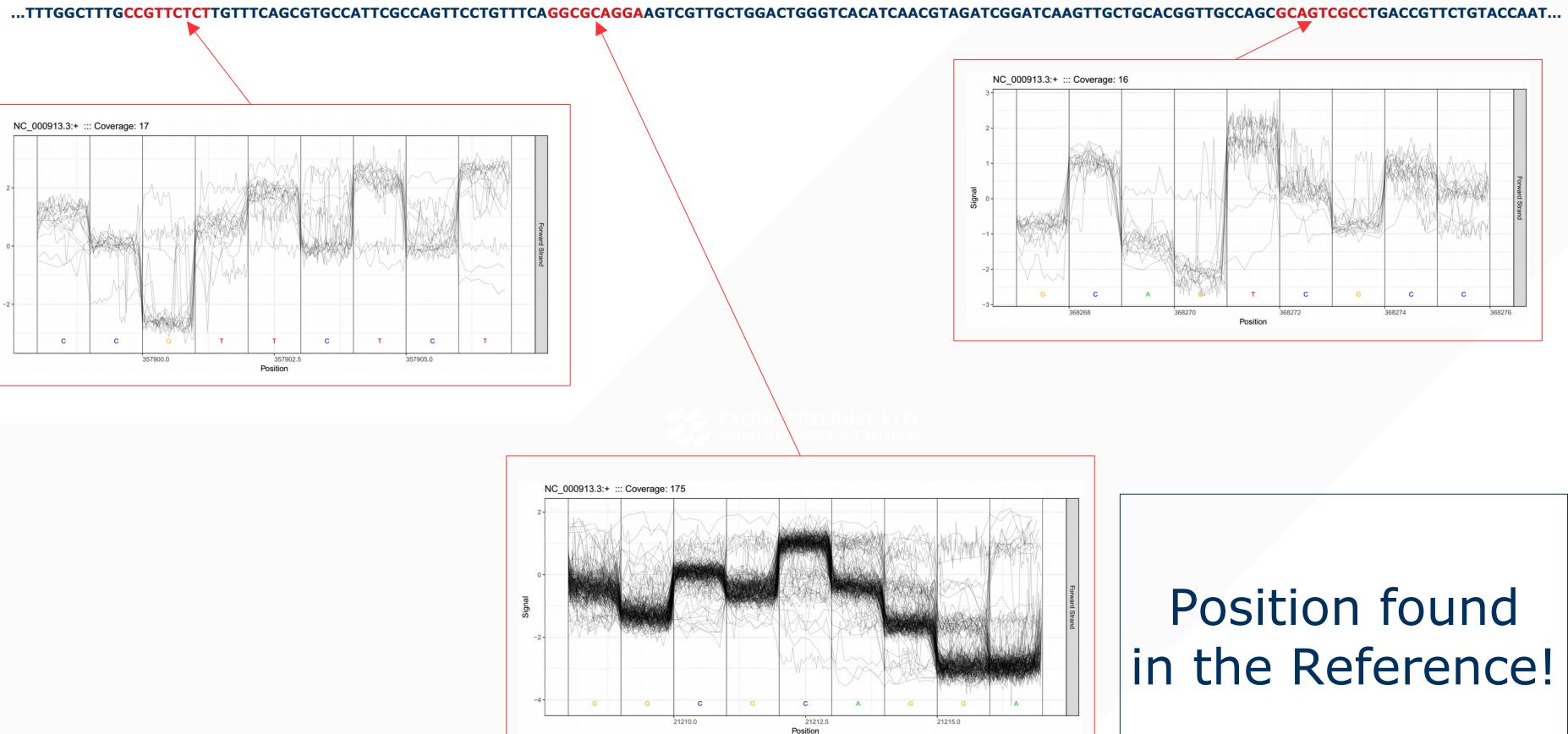
...GGATCTCCAGTGCCTTAC**CCGTTCTCT**GCAGACTACCTGTTGCCAGAAAACAGCGCGTGCCTAAAGCGTCCAGGACATGGAAGCGCGCTCGTGAACCGTATGACGCTGGAAGCGATTGTCACCCAGTCGGACCGCGAC...

# Anchoring Patterns in the Reference Genome Sequence

...TTGGCTTGCCGTTCTCTGTTTCAAGCGTGCCTGCCAGTCCTGTTCAAGGCGCAGGAAGTCGTGCTGGACTGGGTACATCAACGTAGATCGGATCAAGTTGCTGCACGGTTGCCAGCGCAGTCGCCTGACCCTGTACCAAT...



# Anchoring Patterns in the Reference Genome Sequence



15. September 2017 18/23



FACHHOCHSCHULE KIEL  
University of Applied Sciences

# Intermediate Results

- Detection Rate in E.coli Bacteria Reference 80%
  - Good but not usable for bigger references
- Only few discrete Patterns found
- Sample needs to be rather long
- Many Patterns within one sample
- Not usable for bigger references
- Normalization needs to be done
- Time-consuming
- The whole sample is needed
- Not real-time capable

## Future Improvements

- Find more discrete Patterns
  - Samples may be shorter
    - or
  - Reference may be longer
- Use Hidden-Markov-Model for distinguishing Pattern
  - Probability value for the found Pattern
  - Only good matching Patterns may be used
  - Overall accuracy will increase



=> Overall Goal:

Enable this method for the human reference genome

## Acknowledgments

### Thanks to:

- Benedicte Willert & Prof. Dr. Ole Ammerpol  
Institute for Human Genetics,  
University Hospital Schleswig Holstein  
Campus Kiel
- PD Dr. Franz-Josef Müller and Björn Brändl  
Center for Integrative Psychiatry  
University Hospital Schleswig Holstein  
Campus Kiel

?

?

?



FACHHOCHSCHULE KIEL  
University of Applied Sciences

15. September 2017 22/23



**FACHHOCHSCHULE KIEL**  
University of Applied Sciences

**Thank you,  
for your attention.**

15. September 2017 23/23



**FACHHOCHSCHULE KIEL**  
University of Applied Sciences