Hasso Plattner Institut
IT Systems Engineering | Universität Potsdam

Evaluation Exercise I

Dr. Matthieu-P. Schapranow

Data Management for Digital Health
Summer 2017

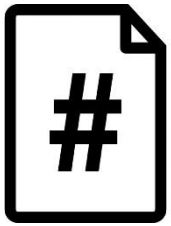
Exercise I
Topics

- Bio recap
- Data acquisition
- Data processing
- Data formats

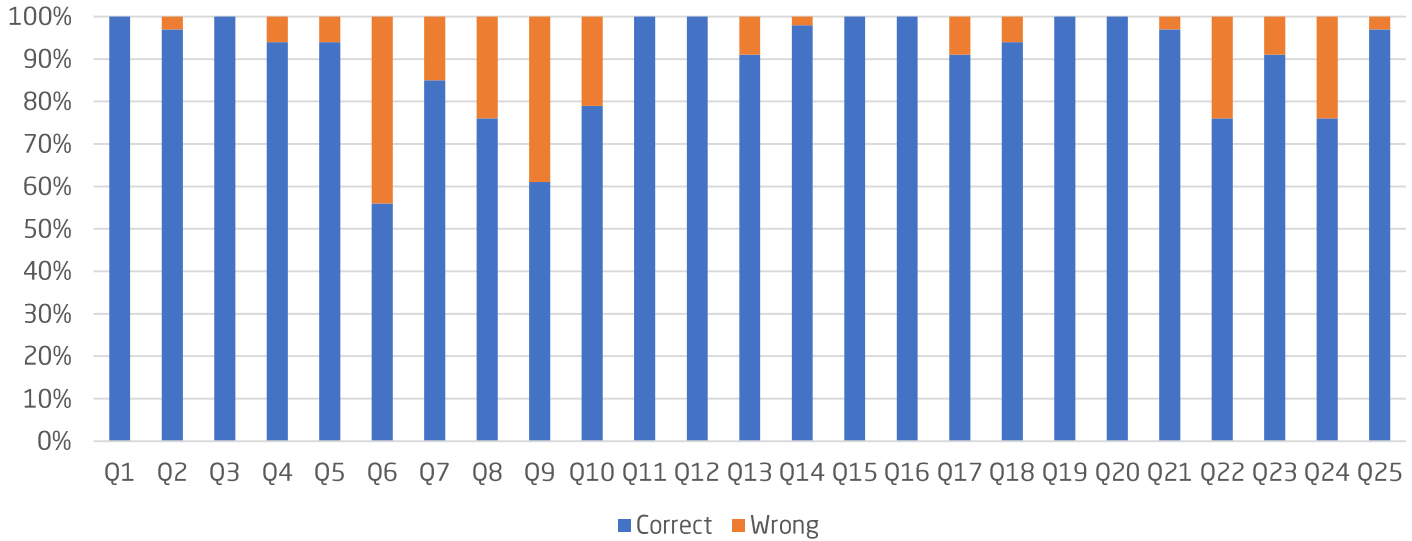
Exercise I
Key Stats

25 Questions 35 Points	33 Students 32 Passed	Average score 31.55 / 90%	Average time 84min



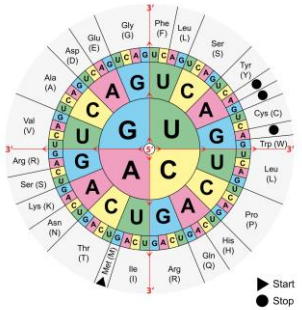



Exercise I
Key Stats



6. The following nucleotide sequence TCA-CCT-GTA-TAG-ATA is given.

- ...and we assume that it is read from left to right and the reading frame starts at the first position. Which of the following statements is correct?
  - **It contains one stop codon**
  - It contains one start codon
  - It contains the amino acid leucine
  - It does not contain the amino acid lysine

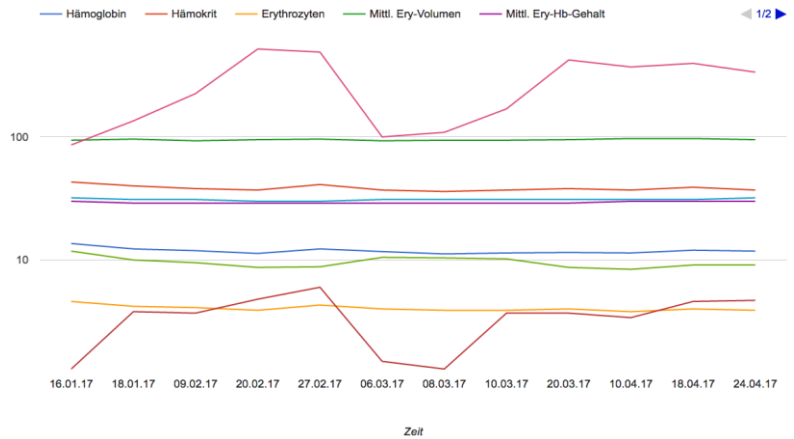


TCA CCT GTA TAG ATA

S P V * I

8. Why are longitudinal data important for health research?

- They are not important at all
- They document equidistant measurements
- They are used as foundation for plotting linear graphs
- **They provide multiple measurements of the same parameter across time**

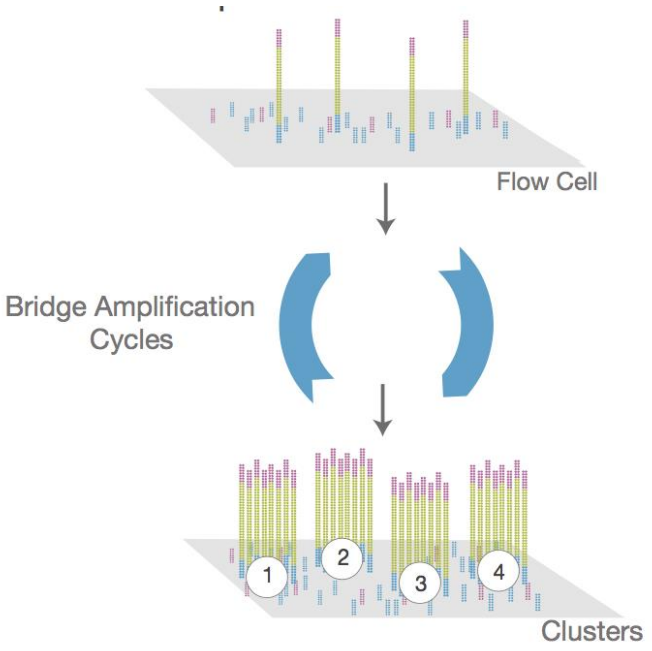


9. In the following, please consider sensor data, such as acquired on intensive care units in hospitals.

- Which of the following statements are correct?
  - **Each measurement has a relatively small data size but they occur at a high frequency**
  - **Recording sensor data over time provides longitudinal data for the specific measurement**
  - Sensor data are generated at a frequency of 100-200 Hz
  - Sensor data are depicted as linear plots

10. Which of the following steps is crucial for current low-cost high-throughput sequencing:

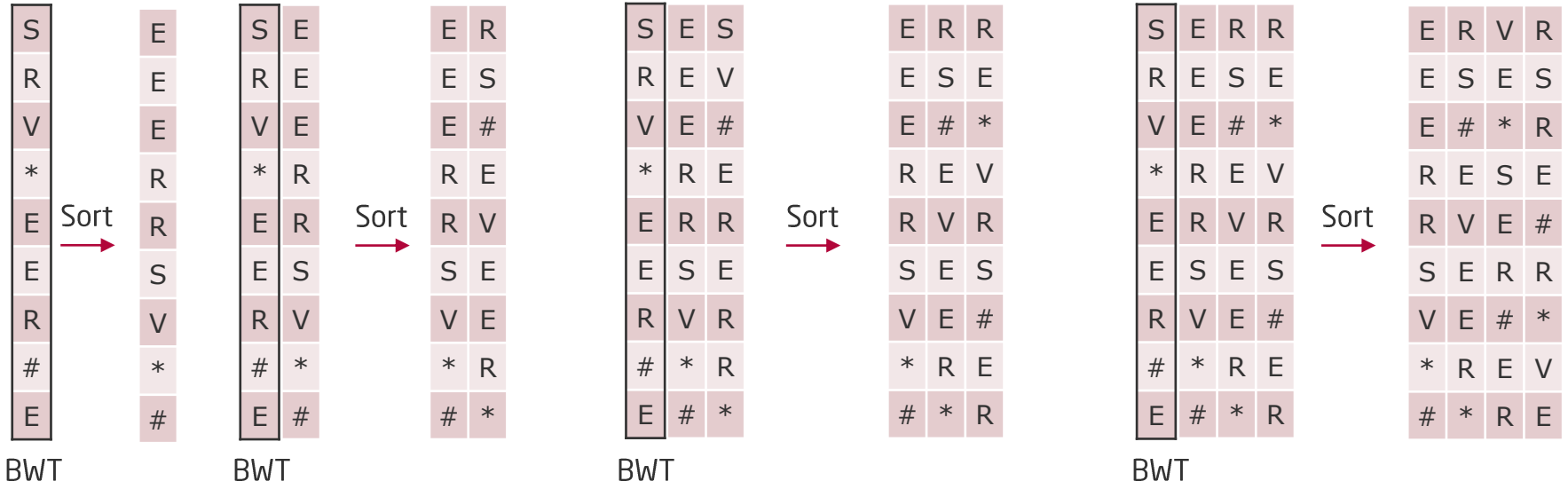
- Nucleotide saturation
- **Amplification of DNA fragments**
- Gene alignment
- Indexing of base pairs



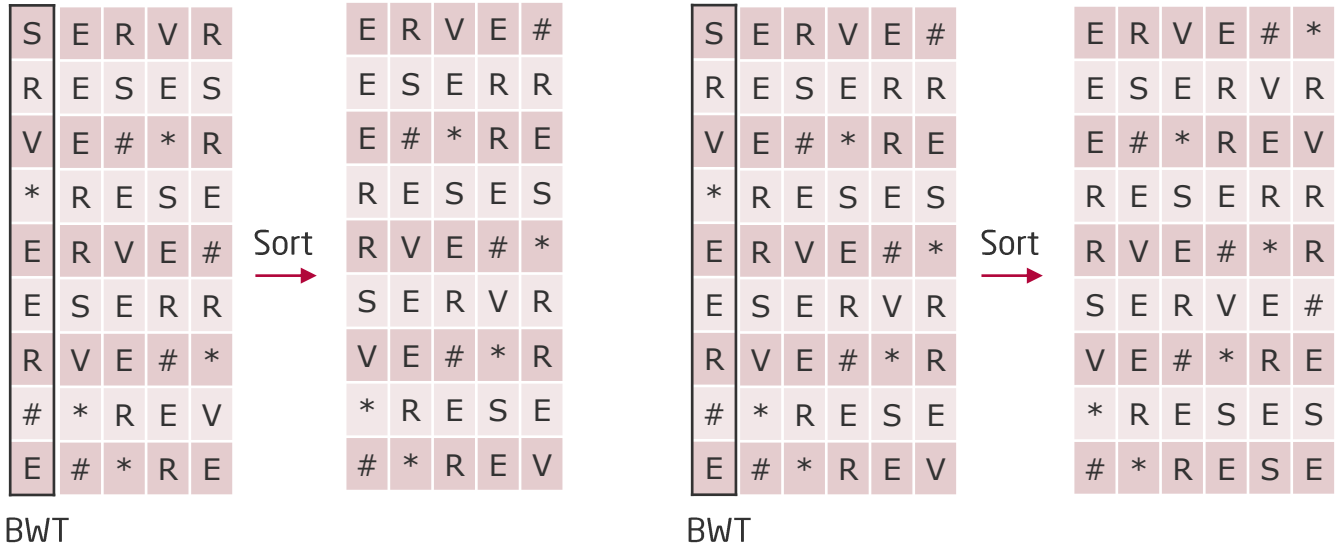
22. Please select all correct statements for Burrows-Wheeler Transform (BWT)

- BWT is a compression schema for nucleotide sequences
- **BWT is a bijective operation**
- **BWT creates from a given input string a reordered output string more suitable for data compression**
- BWT cannot be used for DNA sequences as their alphabet is too limited to benefit from data compression

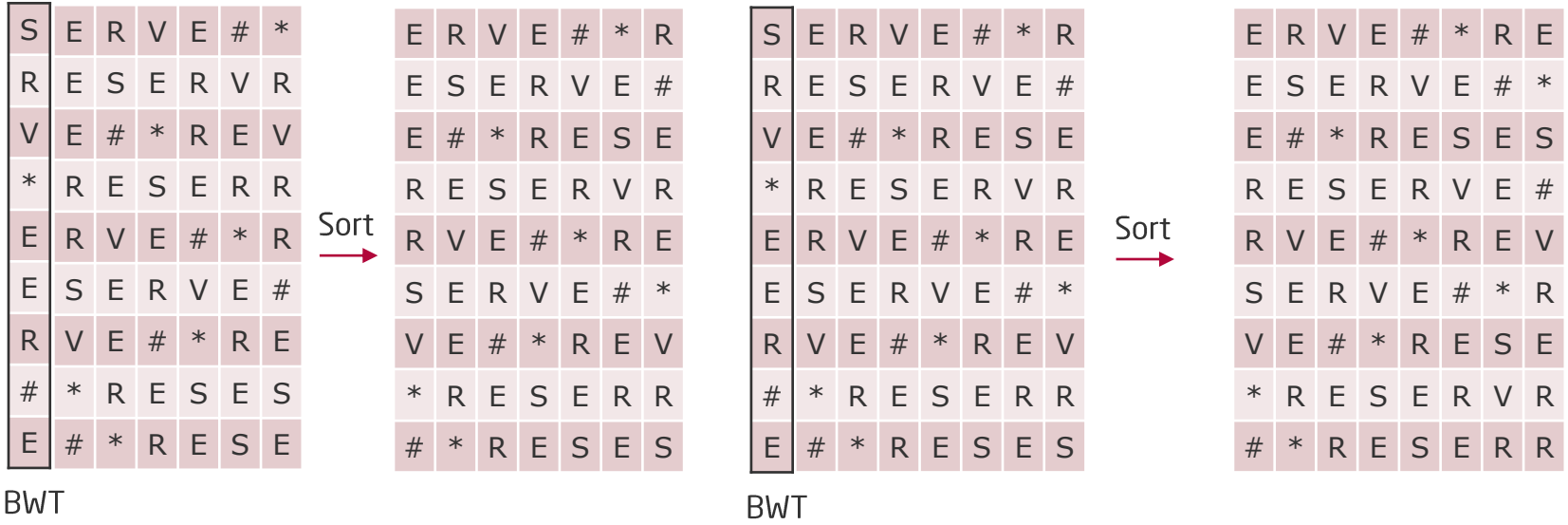
24. Reverse SRV*E²R#E (RLE) applying the Burrows-Wheeler Transform



24. Reverse SRV*E²R#E (RLE) applying the Burrows-Wheeler Transform



24. Reverse SRV*E²R#E (RLE) applying the Burrows-Wheeler Transform



24. Reverse SRV*E²R#E (RLE) applying the Burrows-Wheeler Transform

S	E	R	V	E	#	*	R	E
R	E	S	E	R	V	E	#	*
V	E	#	*	R	E	S	E	S
*	R	E	S	E	R	V	E	#
E	R	V	E	#	*	R	E	V
E	S	E	R	V	E	#	*	R
R	V	E	#	*	R	E	S	E
#	*	R	E	S	E	R	V	R
E	#	*	R	E	S	E	R	R

BWT

Sort →

E	R	V	E	#	*	R	E	V
E	S	E	R	V	E	#	*	R
E	#	*	R	E	S	E	R	R
R	E	S	E	R	V	E	#	*
R	V	E	#	*	R	E	S	E
S	E	R	V	E	#	*	R	E
V	E	#	*	R	E	S	E	S
*	R	E	S	E	R	V	E	#
#	*	R	E	S	E	R	V	R

24. Reverse SRV*E²R#E (RLE) applying the Burrows-Wheeler Transform

- *RESERVE#
- *REVERES#
- *REVERSE#
- *SEVERER#

E	R	V	E	#	*	R	E	V
E	S	E	R	V	E	#	*	R
E	#	*	R	E	S	E	R	R
R	E	S	E	R	V	E	#	*
R	V	E	#	*	R	E	S	E
S	E	R	V	E	#	*	R	E
V	E	#	*	R	E	S	E	S
*	R	E	S	E	R	V	E	#
#	*	R	E	S	E	R	V	R