



Data Management for Digital Health

Revision of Exercise II

Borchert, Dr. Schapranow
Data Management for Digital Health
Winter 2023

Exercise II

Topics

- Medical Use Case Oncology
- Bio Recap
- Genome Data Acquisition and Processing

Evaluation Exercise I

Data Management for
Digital Health, Winter
2023
2

Exercise II

Key Stats

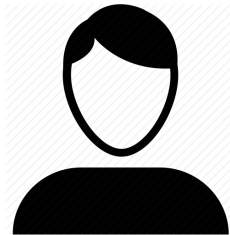
25 Questions
50 Points

27 Students
27 Passed

Average score
40.7 / 81%

Average time
97.3 min

<< 3h

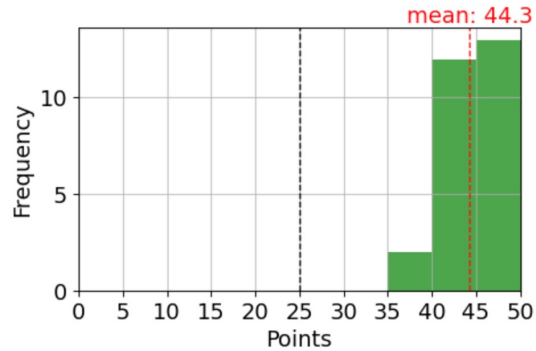


Evaluation Exercise I

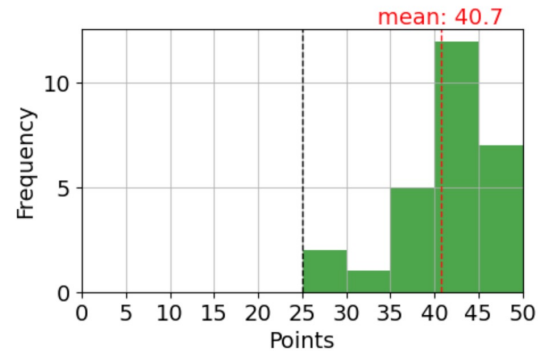
Data Management for
Digital Health, Winter
2023

3

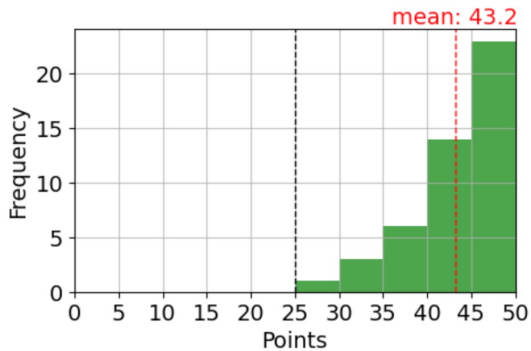
Exercise II Key Stats



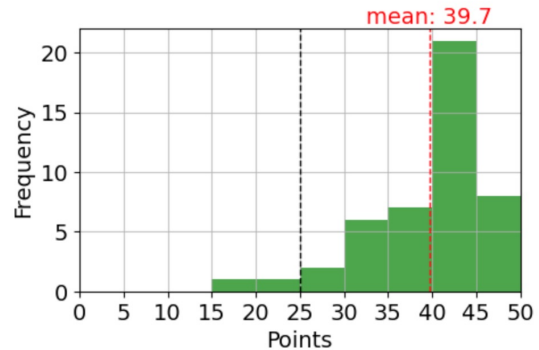
Exercise 1 (2023)



Exercise 2 (2023)



Exercise 1 (2022)



Exercise 2 (2022)

Evaluation Exercise I

Data Management for
Digital Health, Winter
2023

Q13 A genetic variation affects the position 23 and alters nucleobase A to G. Please select all answers describing the resulting effect.

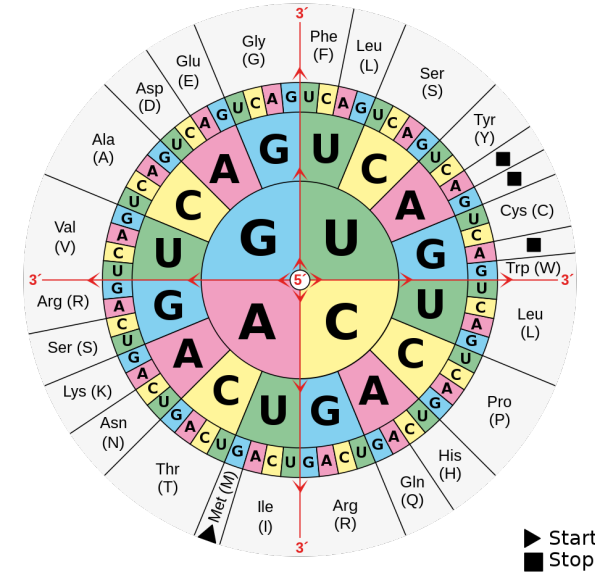
- ✓ The genetic variant is non-functional.
- ✗ The genetic variant is functional.
- ✗ The genetic variant introduces a nonstop mutation.
- ✗ The genetic variant introduces a missense mutation.

Q14 A genetic variation affects position 19 and alters nucleobase A to U. Please select all answers describing the resulting effect.

- ✓ The genetic variant introduces a nonsense mutation.
- ✗ The genetic variant introduces a nonstop mutation.
- ✗ The genetic variant is non-functional.
- ✓ The genetic variant is functional.

Q13: nucleobase A to G at pos 23
 Q14: nucleobase A to U at pos 19

	1-3	4-6	7-9	10-12	13-15	16-18	19-21	22-24
	*	W	I	N	T	E	R	.
	AUG	UGG	AUC	AAC	ACG	GAG	AGA	UAA
Q13	AUG	UGG	AUC	AAC	ACG	GAG	AGA	UGA
Q13	*	W	I	N	T	E	R	.
Q14	AUG	UGG	AUC	AAC	ACG	GAG	UGA	UAA
Q14	*	W	I	N	T	E	.	.



Evaluation Exercise I

Data Management for
Digital Health, Winter
2023
7

Point Mutations / Gene Mutations

- **Single Nucleotide Polymorphism (SNP)** := Affects a single locus on a gene, e.g. substitution of a single base
- **In/Del** := Insertion/Deletion of an arbitrary number of bases resulting in a frame shift
- **Non-functional** := No impact on products created from the affected genetic code, e.g. compensated through amino acids redundancy
- **Functional** := Impact on products built from affected genetic code, e.g.:



Size (before vs. after)	Type	Impact
=	Missense	Changes triplet, i.e. another amino acid chain is synthesized
>	Nonsense	Converts existing triplet to stop codon (+STOP)
<	Nonstop	Converts existing stop codon to another triplet (-STOP)

Medical Use Case Oncology

Data Management for
Digital Health, Winter
2023
8

Q15 In the following, please incorporate your observations from our DIY Build Your Own DNA Workshop to select all correct statements.

- ✓ After combining individual pieces of DNA assembled in groups, we were able to read the sentence “We learn Digital Health”.
- ✗ Each group created the same sequence of DNA and applied different kinds of mutations to it to make it unique.
- ✗ “Make biology great again” was encoded in the given DNA strand.
- ✗ Each group created a special chunk of DNA and applied the same type of mutation to it to make it unique.

Frequently missed

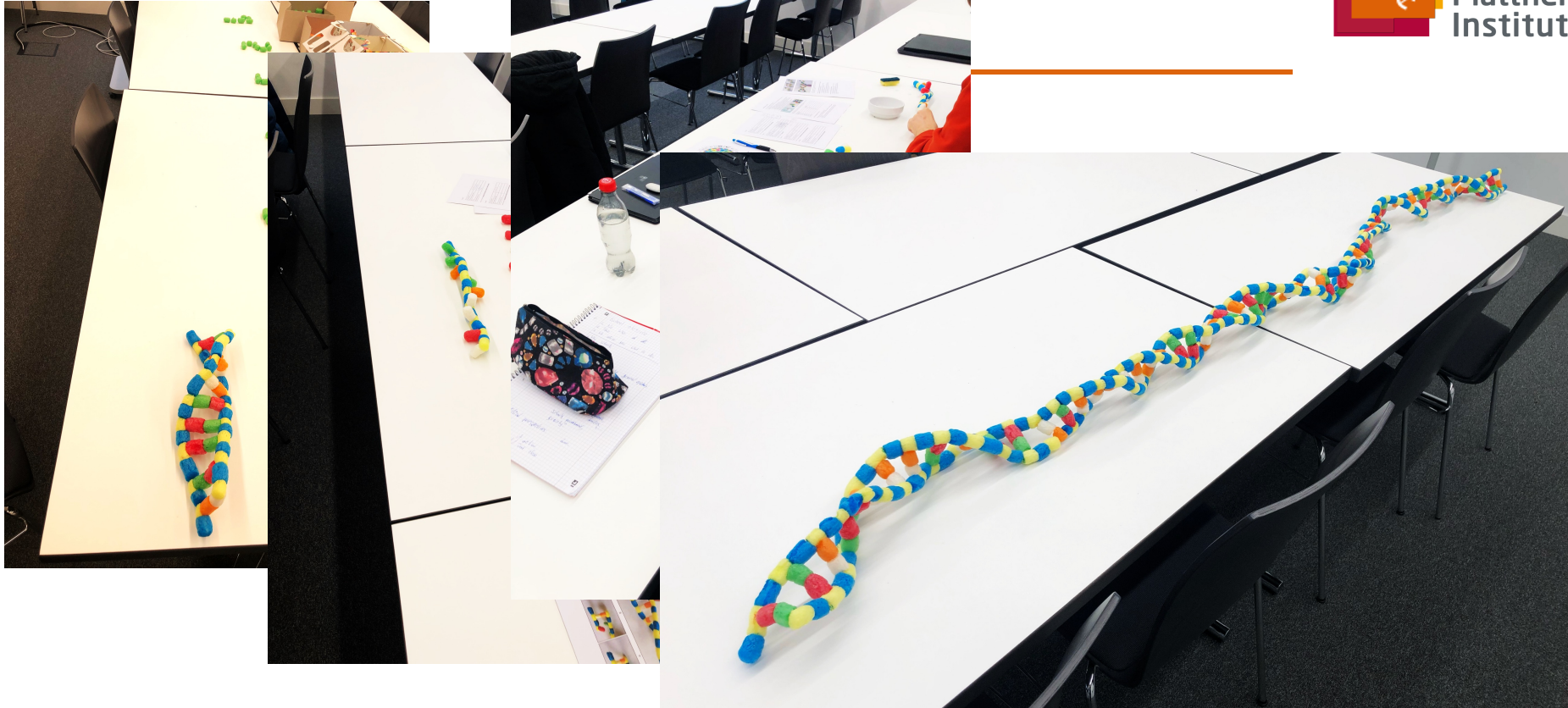
Frequent incorrect answer

Evaluation Exercise I

Data Management for
Digital Health, Winter
2023

9

DIY Workshop



Q18: ... please make use of the Needleman-Wunsch alignment algorithm [...] for the given sequence ATTA and the reference GATTACA

✗ ■ 2

✓ ■ -4

✗ ■ 0

✗ ■ -2

	-	G	A	T	T	A	C	A
-	0	← -4	← -8	← -12	← -16	← -20	← -24	← -28
A	↑ -4	↖ -2	↖ -2	← -6	← -10	← ↖ -14	← -18	← ↖ -22
T	↑ -8	↑ ↖ -6	↖ -4	↖ 0	← ↖ -4	-8	← -12	← -16
T	↑ -12	↑ ↖ -10	↑ ↖ -8	↖ -2	↖ 2	← -2	← -6	← -10
A	↑ -16	↑ ↖ -14	↖ -8	↑ -6	↑ -2	↖ 4	← 0	← ↖ -4

Gap cost function $gap() := -4$


Weight function $w(a,b) :=$
 $\{ 2 \text{ if } a=b \text{ else } -2. \}$

Evaluation Exercise I

Data Management for Digital Health, Winter 2023
11

Q19: The best global alignment score of the given sequences contains:

 Three mismatches, no gaps

 One gap, two mismatches

 Two gaps, one mismatch

 Three gaps, no mismatches

	-	G	A	T	T	A	C	A
-	0	← -4	← -8	← -12	← -16	← -20	← -24	← -28
A	↑ -4	↖ -2	↖ -2	← -6	← -10	← ↖ -14	← -18	← ↖ -22
T	↑ -8	↑ ↖ -6	↖ -4	↖ 0	← ↖ -4	-8	← -12	← -16
T	↑ -12	↑ ↖ -10	↑ ↖ -8	↖ -2	↖ 2	← -2	← -6	← -10
A	↑ -16	↑ ↖ -14	↖ -8	↑ -6	↑ -2	↖ 4	← 0	← ↖ -4

```

G A T T A C A
_ A T T A _ _
    
```

Evaluation Exercise I

Data Management for Digital Health, Winter 2023
12

Q20: [...] In the following, please consider the given string "L*BMA³#A", which was BWT+RLE-encoded using the assumptions *=start, #=end, num(*) < num(#). Please apply the inverse BWT function BWT⁻¹ to the given string.

- ✓ ■ The original string contains the letter 'A' four times.
- ✗ ■ The original string contains three consecutive letters 'A'.
- ✗ ■ The original string contains 'B' and 'L', whereas 'B' occurs first.
- ✓ ■ The original string contains 'L' and 'M', whereas 'L' occurs first.

Read	Sorted	-	BWT
4.	A ₁	<	L
2.	A ₂	<	*
6.	A ₃	<	B
8.	A ₄	<	M
5.	B	<	A ₁
3.	L	<	A ₂
7.	M	<	A ₃
1.	*	<	#
9.	#	<	A ₄

1. Add indices for multiple occurrences, i.e. A₁..A₄, in the same order in both columns
2. Lookup '#' in the BWT column, because we know it is the end of our string.
3. Add the corresponding value from sorted column to output, i.e. *
4. Lookup * in the BWT column
5. Add the corresponding value from sorted column to output, i.e. A₂
6. ...

Evaluation Exercise I

Q24: Please select data management challenges for precision oncology as discussed in class.

- ✓ ■ Genome data processing requires cooperation between multiple disciplines and user roles, e.g., clinical researcher, lab assistant, bioinformatician, etc.
- ✗ ■ The Genome Data Processing and Modeling Notation (GDPMN) allows the definition and exchange of data management processes for precision oncology in a standardized format.
- ✗ ■ The use of centralized cloud solutions (*by third-party vendors*), e.g., for processing of genetic data and patient data, facilitates the adoption of precision oncology.
- ✓ ■ Processing of genetic data is time- and resource-intensive.

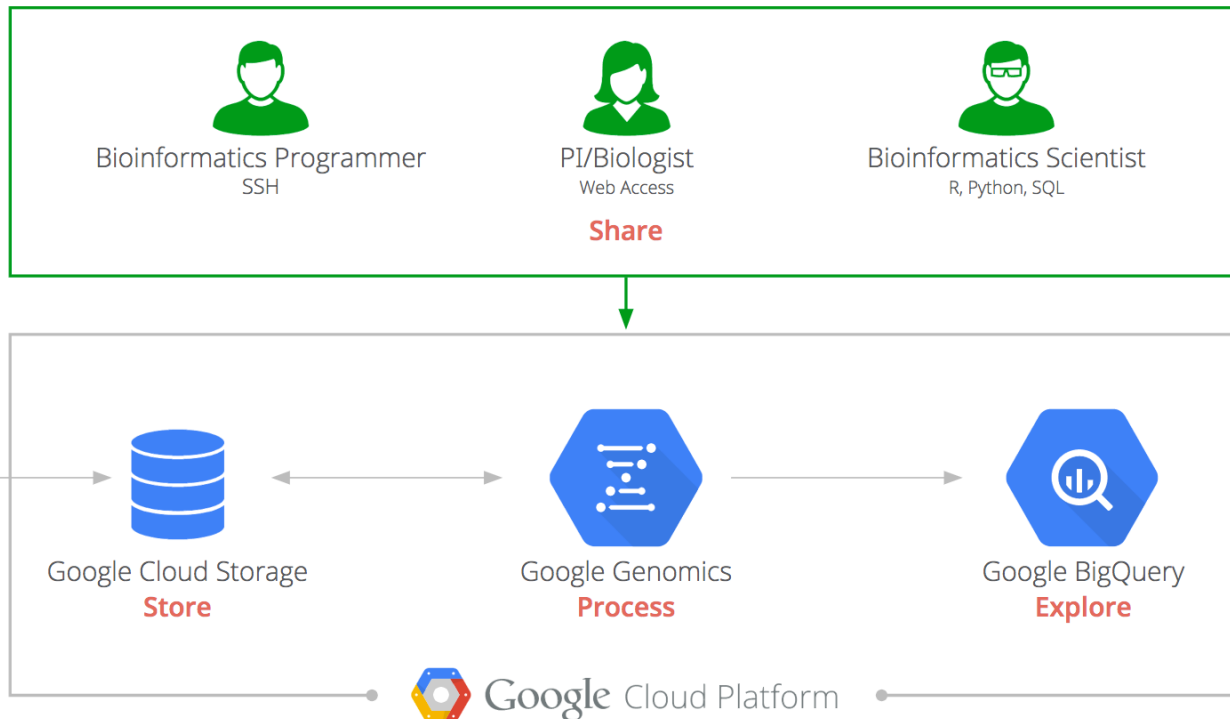
Frequently missed

Frequent incorrect answer

Evaluation Exercise I

Data Management for
Digital Health, Winter
2023
16

■ Integration of existing Google services to genome data processing



**Data Management for
Precision Oncology**

Data Management for
Digital Health, Winter
2023
17