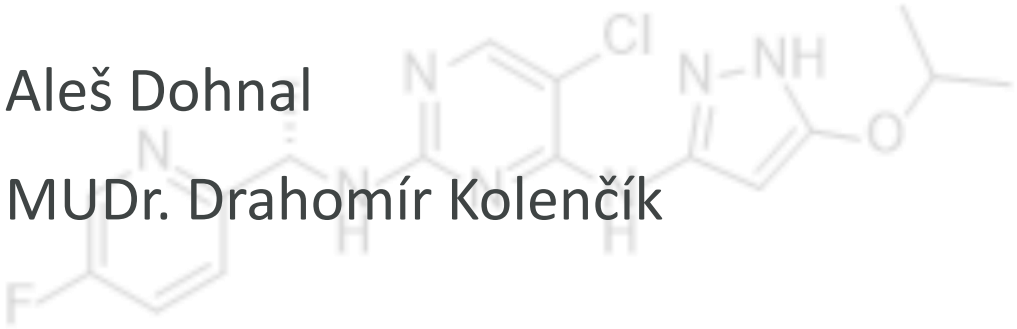
AstraZeneca

European Healthcare Hackathon 2023

Icebreaker AZ23

Aleš Dohnal

MUDr. Drahomír Kolenčík



Pathology is bridge between science and medicine. Can you break the ice between complex genomic data and cancer researchers?

Find the courage to take up a challenge that aims to tackle the issue of efficient processing and use of sequential data. You can fundamentally change current practice.

Why?

PATHOLOGY IS THE MOTOR THAT DRIVES HEALTHCARE TO UNDERSTAND DISEASES. WHILE IT DOES THE JOB VIA THE SAME METHODS THAT IT HAS BEEN USING FOR THE LAST 150 YEARS, IT'S TIME TO CHANGE.

DIGITAL TECHNOLOGIES COULD PUSH THE FIELD INTO BECOMING MORE EFFICIENT AND MORE SCALABLE.

IT CAN TRANSFORM THE JOB OF PATHOLOGISTS INTO A MORE CREATIVE AND DATA-DRIVEN PROFESSION WHILE ALLOWING PATIENTS TO RECEIVE DIAGNOSES FASTER AND MORE ACCURATELY.

https://medicalfuturist.com/digital-future-pathology/

What do you think is pathology?

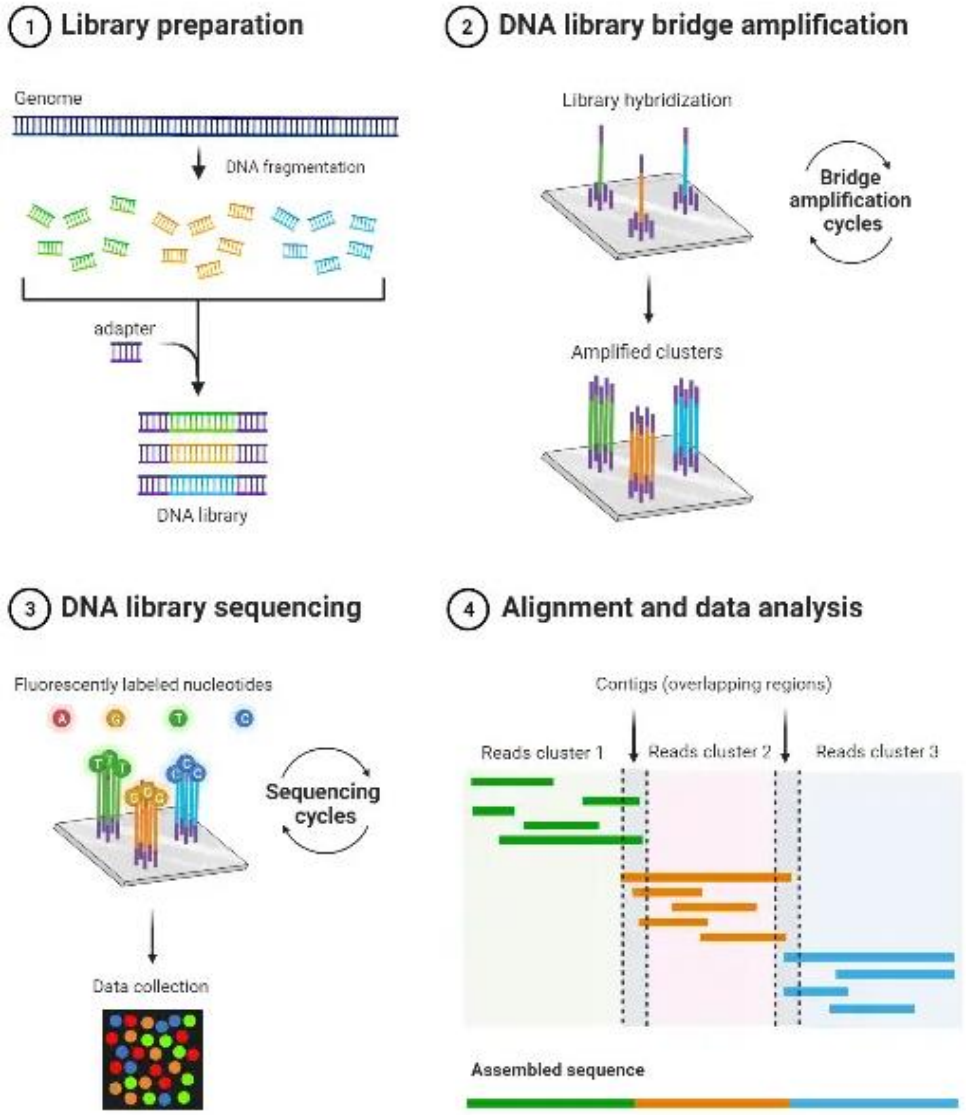
What is pathology?

Pathology is the medical discipline that provides diagnostic information to patients and clinicians. It impacts nearly all aspects of patient care, from diagnosing cancer to managing chronic diseases through accurate laboratory testing.

Pathologists work closely with surgeons, radiologists, and oncologists. Pathologists can sub-specialize in different areas, such as gastroenterology, gynecologic pathology, blood diseases, clotting disorders, microbiology, lung and breast cancers, and more. For every sub-specialty in medicine or surgery, there is a pathologist counterpart, helping to make the correct diagnosis and guide the care of the patient.

Next-Generation Sequencing in Diagnostic Pathology

The pathologist will usually evaluate tissue morphologically and expression of biomarkers. Recent developments in **sequencing technology** means that DNA and RNA can be now use for the diagnostic purposes as well. These new technologies, collectively known as **next-generation sequencing** (NGS), generate **huge amounts of data** which can be used to support patient management



Challenge define with Institute of Pathology 1st Faculty of Medicine Charles University and General University Hospital

Build solid but flexible data model

Find a way how to efficiently combine, maintain and share information about test , performed sequential tests and patients.

- ➢ Sample files available
- ➢ Key information marked

Index	Chrom	Pos	Coverage	Ref	Mutant Allele Frequency	Ins%	Del%	Overall Score	Mutation Call:Genomic	Mutation Call:Relative To CDS	Mutation Call:HGVS Coding	Function	Gene	Read Balance(Percentage)	SNP db_xref (dbsnp 135)	Amino Acid Change
46	1	155870189	2533	G	64,51	0	0	26,6	G>CG	c.701C>C	NM_0012	Nonsense	RIT1	0,99		p.S234XS
951	18	48581309	1011	G	92,48	0	0	23,2	G>GT	c.613G>G	NM_0053	Nonsense	SMAD4	0,99		p.E205EX
1092	22	24176339	526	G	43,73	0	0	21,2	G>AG	c.1130G>A	NM_0030	Missense	SMARCB1	0,97		p.R377HR
31	1	118166385	955	G	96,13	0	0,21	23,4	G>A	c.895G>A	NM_0177	Missense	FAM46C	1	rs14547178	p.E299K
766	13	110436782	236	T	47,88	0	0	18,7	T>GT	c.1619A>A	NM_0037	Missense	IRS2	0,97		p.Y540SY
936	17	76219604	1393	G	25,63	0	0	24,8	G>GT	c.467G>G	NM_0010	Missense	BIRC5	0,98		
50	1	156830779	240	G	25	0	0	18,8	G>AG	c.123-336	NM_0010	Noncoding	NTRK1	0,95	rs100721	
62	1	158587390	966	C	35,51	0	0	22,8	C>CT	c.6549-12	NM_0031	Noncoding	SPTA1	0,94	rs857716	
845	16	2121869	386	C	51,55	0	0	19,2	C>CT	c.2031C>C	NM_0005	Synonymo	TSC2	0,92	rs455172	
868	16	89805915	791	G	51,58	0	0	23,1	G>AG	c.3981C>C	NM_0001	Synonymo	FANCA	0,99	rs141278	
193	2	234668880	347	A	34,87	34,87	2,31	16	insTA	c.856-678	c.856-678	Noncoding	UGT1A	0,73	rs675742	
17	1	27023381	99	G	5,05	0	0	15,1	G>CG	c.487G>C	NM_0060	Missense	ARID1A	0,53		
18	1	27023450	190	A	5,26	0	0	16,5	A>AC	c.556A>A	NM_0060	Missense	ARID1A	0,54		
22	1	27023942	34	T	11,76	0	0	12,1	T>GT	c.1048T>G	NM_0060	Missense	ARID1A	1		
166	2	173420994	317	A	5,36	0	0	18,7	A>AC	c.116A>A	NM_0026	Missense	PDK1	0,87		
353	5	56111806	120	T	6,67	0	0	16,5	T>GT	c.406T>G	NM_0059	Missense	MAP3K1	0,72		
440	6	86159885	148	G	8,11	0	0,68	8,1	G>CG	c.28G>CG	NM_0025	Missense	NT5E	0,64		
463	6	157099713	112	T	9,82	0	0	10,1	T>GT	c.650T>G	NM_0207	Missense	ARID1B	0,7		
466	6	157100026	57	A	5,26	0	5,26	13,4	delA	c.963delA	NM_0207	Frameshift	ARID1B	0,42		
475	6	157100431	32	G	9,38	0	9,38	12	delGG	c.1368_13	NM_0207	Frameshift	ARID1B	0,83		

příjem LMP	uzavření LMP	doba odezvy	dg	VFN - onkologie	kódy	Biopsie
22.12.2021	10.01.2022	12	DG - prediktivní - jiné	C787		B20509/21
22.12.2021	10.01.2022	12	DG - prediktivní - prs	C509		B20584/21
23.12.2021	10.01.2022	11	DG - prediktivní - jiné	C23		B20506/21
27.12.2021	10.01.2022	10	DG - prediktivní - jiné	C787		B20508/21
27.12.2021	10.01.2022	10	DG - prediktivní - CRC	C189		B20507/21
30.12.2021	10.01.2022	7	DG - prediktivní - jiné	C169 žaludek		B20736/21
31.12.2021	11.01.2022	7	BRCA HGSC	C56		B20733/21
03.01.2022	11.01.2022	7	DG - prediktivní - jiné	C787		B20780/21
03.01.2022	10.01.2022	6	DG - prediktivní - jiné	C119		B20691/21
04.01.2022	11.01.2022	6	BRCA HGSC	C56		B20797/21
04.01.2022	11.01.2022	6	dif dg - atypický Spitz névus	C229		B00054/22
08.08.2023	24.08.2023	13	DG - prediktivní - jiné	C800	87807+87808	B14211/23
09.08.2023	22.08.2023	10	DG - prediktivní - jiné	C800	87807+87808	B14283/23
10.08.2023	22.08.2023	9	DG - prediktivní - CRC	C184	87805+87806	B14291/23
11.08.2023	24.08.2023	10	DG - prediktivní - jiné	C800	87807+87808	B14053/23

Číslo biopsie	Název bloku	Pozn.	% nádorových buněk	DNA konc. po 1.PCR	DNA průměrné pokrytí	DNA CNV	DNA TMB	DNA MSI	DNA mutace	
14211/23	DG-14013-23-1A	Poncar	90	176	366		5	MSS	TP53	3mi
14291/23	DG-13217-23-1G		70	179	431	Del.: FGFR1; Dupl.: CDKN2A	3	MSS	TP53, KRAS, BRAF, TSC2, UGT1A1	4.5
14340/23	DG-13129-23-5A-8A		50	124	429	Del.: TP53	4	MSS	NRAS, JUN, WNK1, APC, SYK	5.2
14053/23	DG-14328-23-1A	Poncar	80	169	421		5	MSS		8.2
14452/23	DG-C2169-23		90	124	595	nelze hodnotit	2	MSS	TP53, RB1	6mi
14518/23	DG-16943-17-2C		80	31	563	duplikace genu NTRK1, FGFR3, ST	17	MSS	APC, TP53, PIK3CA, ARID1A	1.8
14436/23	DG-43309-23-B4		50	8	572	nelze	9	MSS	APC, TP53, AR, WRN	1.3
14398/23	BRCA-7675-23-3		80	78,8	462	BRCA1/2 normal	4	MSS	BRCA1/2 WT	NA
14611/23	DG-4296-23-3	vyizolovaná DNA	90	86	530	nic	5	MSS	ARID1A	NA
14338/23	DG-14338-23-plasma-UMI	Společně s FastGen	plasma	19,1	132	nelze, nízké pokrytí a cfDNA	nelze	nelze	Nic	NA
14283/23	DG-1-122764-21-j-UMI		60	37,7	1213	nelze spolehlivě	3	MSS	žádná class 4/5	6.2

Known obstacles

Data can be generated by different softwares

Different organizations keep data in different structure

Medical development brings the need to verify results against new findings

NGS data needs to better connected with data in clinical setting

Enable

- Manual mapping of columns for loads
- Access to aggregated overview with possibility to dril to individual tests
- Annotation and tagging/color highlighting of test results
- Customizable column naming for exports
- Connection with additional datasources

Usefull links

https://www.cbioportal.org/

https://www.ncbi.nlm.nih.gov/clinvar/

https://cancer.sanger.ac.uk/cosmic

https://www.clinicaltrials.gov/

https://www.citeline.com/

… let’s start the change together today.

Do not use layouts in this section appearing after this point.