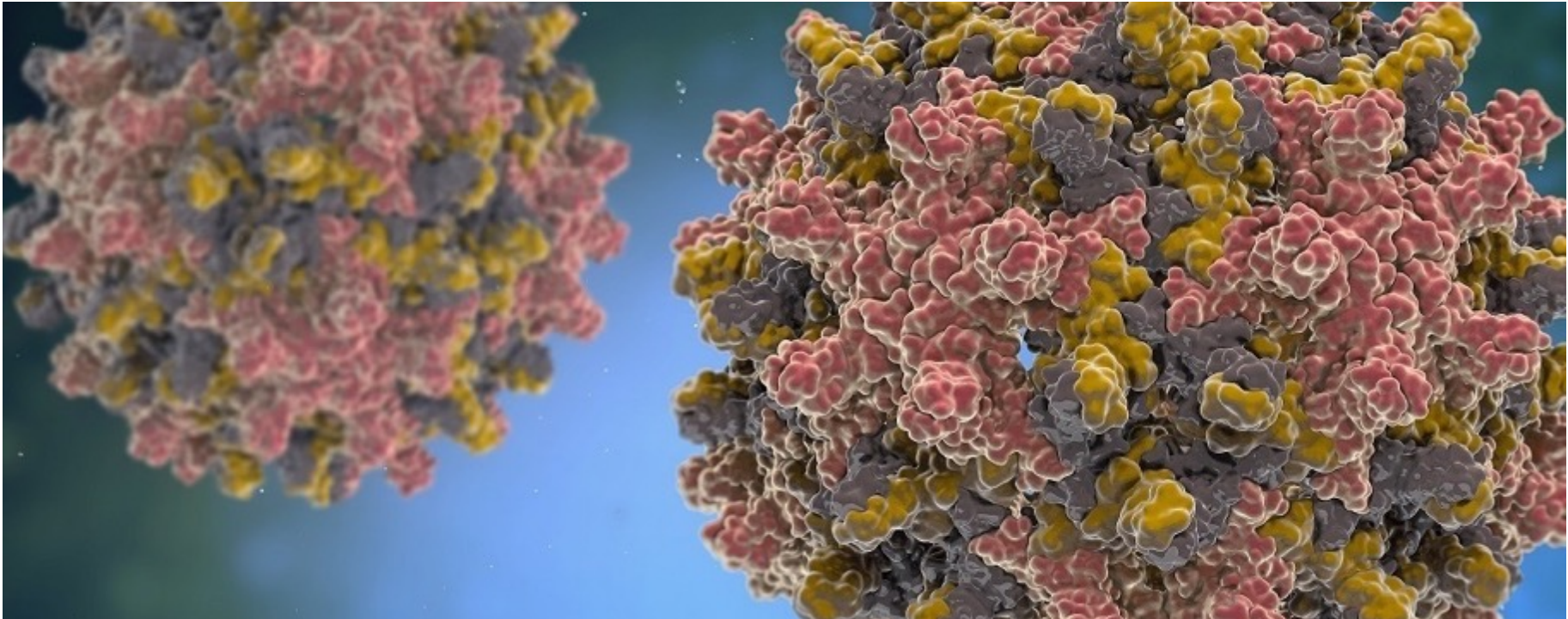




R-Shiny Application for Exploring Deep Sequencing Data

PHUSE EU Connect 2022

Ilse Augustyns, Technical Lead Methodology & Innovation
Annemie Van Dijck, Principal Statistical Programming Lead

Janssen R&D
NOV2022

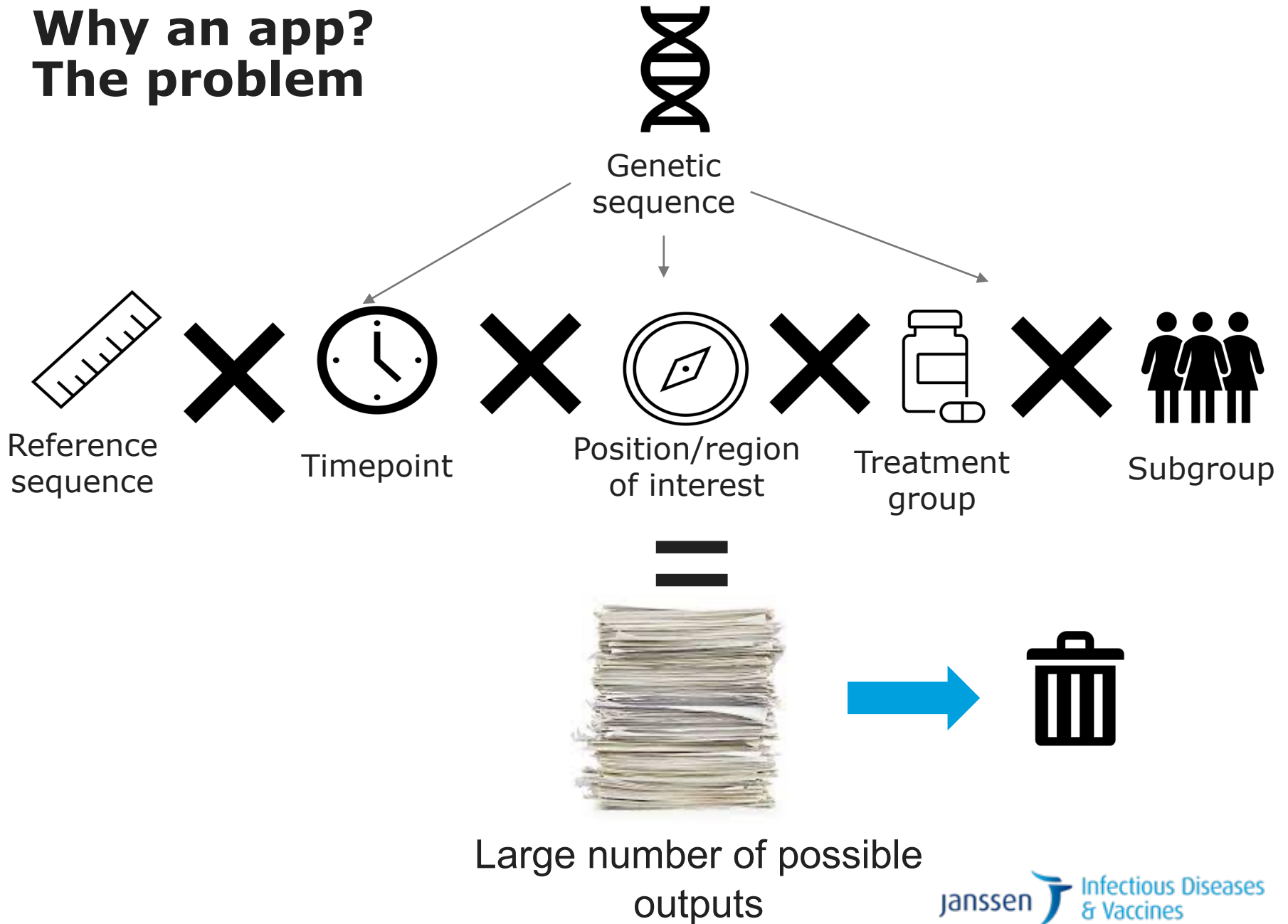
Agenda

1	Deep Sequencing	Annemie Van Dijck
2	The Exploration App	Annemie Van Dijck
3	The Technical Side	Ilse Augustyns
4	Q&A	

Why Deep Sequencing?

- Anti-viral treatments act on different target regions
- Reveal genetic variations in the target genes
 - ➡ better or worse response to these treatments

Why an app? The problem



Why an app? The solution



The input: ADPF dataset

- Based on the SDTM PF + SUPPPF dataset
- Added:
 - Add variables (merge with covariates, population flags and treatment groups)
 - Add timing variables (ady, avisit(n) and period)
 - Add genotype (can be derived from reference used)
- One record per subject, visit, position and variation
- Derivation about the observed variable and component:
 - Observed variant
 - Observed component (all/above 15% frequency)
 - For post-baseline visit: emerging and enriched variants

	POSITION	ANVALC	OVAR	OCOMP	OCOMP15	BCOMP	OFREQ
89	184		Q184K	Q184K	Q184K	Q184K	100
90	184		Q184K	Q184K/Q/STOP	Q184K	Q184K/Q/STOP	77.373
91	184		Q184Q	Q184K/Q/STOP	Q184K	Q184K/Q/STOP	11.636
92	184		Q184STOP	Q184K/Q/STOP	Q184K	Q184K/Q/STOP	10.841
93	186		STOP186Q	STOP186STOP/Q		STOP186STOP/Q	1.76
94	186		STOP186STOP	STOP186STOP/Q		STOP186STOP/Q	97.848

Preparation steps

- Create script: study specific data  general format
- Load data to server
- Only action needed when new data is available
- Automated in the future?

100 nm

The exploration app

Transmission electron micrograph of Staphylococcus_bacteria

PhUSE 2022 NGS



 Tables



>> Single Region

 Exploration

Baseline/Emerging/Enriching mutations at a specific timepoint

Treatment Response Tables

Input specifications

Select Population

Intent-to-Treat



Select Type of Analysis

Baseline



Input specifications

Select Population

Intent-to-Treat



Select Type of Analysis

Baseline



Select Type of Presentation

count/Denom (%)



Select Reference Sequence

Universal



Select subgroup variable

Genotype



Label for Subgroup on output

Genotype

Select Subgroup Level

All



Select Level

AMINO ACID



Select Analysis Type

Position (only)



Select Gene region

R1



Select list of positions of interest

15 Region 1



Generate table

Table: Mutations at Specific Timepoint

Input specifications

Select Population

Intent-to-Treat

Select Type of Analysis

Baseline

Select Type of Presentation

count/Denom (%)

Select Reference Sequence

Universal

Select subgroup variable

Genotype

Label for Subgroup on output

Genotype

Select Subgroup Level

Generate table



Download report

TVIPMAAUBLPO103-ITT: Number and Proportion (%) of Subjects With Polymorphisms on Amino Acid Level Versus the Universal Reference Sequence at Baseline (BLSEQ) by Subgroups (Genotype) by Position Considering 15 Region 1 Positions of Interest; Intent-to-Treat (Study PhUSE 2022 Demo)

	Treatment 1	Treatment 2	Treatment 3
Analysis set: Intent-to-Treat	45	48	93
Genotype : All Subjects	45	48	93
Subjects with Sequencing Data (R1)	15	17	32
no baseline polymorphism at positions of interest	12/15 (80.0%)	12/17 (70.6%)	25/32 (78.1%)
1 or More Baseline Polymorphism at positions of interest	3/15 (20.0%)	5/17 (29.4%)	7/32 (21.9%)
24	1/15 (6.7%)	1/17 (5.9%)	4/32 (12.5%)
25	0	0	1/32 (3.1%)
29	1/15 (6.7%)	1/17 (5.9%)	0
30	0	1/17 (5.9%)	0
33	1/15 (6.7%)	0	0
105	1/15 (6.7%)	3/17 (17.6%)	4/32 (12.5%)
109	0	1/17 (5.9%)	0
Genotype : Genotype A	5	2	9
Subjects with Sequencing Data (R1)	2	1	6
no baseline polymorphism at positions of interest	2/2 (100.0%)	0	5/6 (83.3%)
1 or More Baseline Polymorphism at positions of interest	0	1/1 (100.0%)	1/6 (16.7%)
24	0	0	1/6 (16.7%)
105	0	1/1 (100.0%)	1/6 (16.7%)
109	0	0	0

Table: Mutations at Specific Timepoint

Label for Subgroup on output

Genotype

Select Subgroup Level

All

Select Level

NUCLEOTIDE

Select Analysis Type

Position (only)

Select Gene region

R4

Select list of positions of interest

20 Region 4

Generate table

TVIPMNTUBLPO301-ITT: Number and Proportion (%) of Subjects With Polymorphisms on Nucleotide Level Versus the Universal Reference Sequence at Baseline (BLSEQ) by Subgroups (Genotype) by Position Considering 20 Region 4 Positions of Interest; Intent-to-Treat (Study PhUSE 2022 Demo)

	Treatment 1	Treatment 2	Treatment 3
Analysis set: Intent-to-Treat	45	48	93
Genotype : All Subjects	45	48	93
Subjects with Sequencing Data (R4)	15	17	32
no baseline polymorphism at positions of interest	14/15 (93.3%)	15/17 (88.2%)	29/32 (90.6%)
1 or More Baseline Polymorphism at positions of interest	1/15 (6.7%)	2/17 (11.8%)	3/32 (9.4%)
261	0	0	0
273	1/15 (6.7%)	2/17 (11.8%)	3/32 (9.4%)
276	0	0	1/32 (3.1%)
280	0	0	0
Genotype : Genotype A	5	2	9
Subjects with Sequencing Data (R4)	2	1	6
no baseline polymorphism at positions of interest	2/2 (100.0%)	0	6/6 (100.0%)
1 or More Baseline Polymorphism at positions of interest	0	1/1 (100.0%)	0
261	0	0	0
273	0	1/1 (100.0%)	0
280	0	0	0
Genotype : Genotype B	3	2	4
Subjects with Sequencing Data (R4)	2	2	3
no baseline polymorphism at positions of interest	1/2 (50.0%)	1/2 (50.0%)	3/3 (100.0%)
1 or More Baseline Polymorphism at positions of interest	1/2 (50.0%)	1/2 (50.0%)	0
273	1/2 (50.0%)	1/2 (50.0%)	0



Tables



>> Single Region



Exploration

Baseline/Emerging/Enriching mutations at a specific timepoint

Treatment Response Tables

Input specifications

Select Population

Intent-to-Treat



Select subgroup variable

Genotype



Summary

Details

Treatment Response Tables: Summary

Input specifications

Select Population

Intent-to-Treat ▼

Select subgroup variable

Genotype ▼

Label for Subgroup on output

Genotype

Select Subgroup Level

All ▼

Select Level

AMINO ACID ▼

Select Analysis Type

Position (details with mixture ▼

Summary

Details

Download report (one html file including all positions)

Summary Table

Position	Letter Abbreviation	Subjects with baseline poly?	Note
23	A	No subjects with baseline poly at position 23	
24	B	At least one subject with baseline poly at position 24	
25	C	At least one subject with baseline poly at position 25	
29	D	At least one subject with baseline poly at position 29	
30	E	At least one subject with baseline poly at position 30	
33	F	At least one subject with baseline poly at position 33	
37	G	No subjects with baseline poly at position 37	Table not produced as the result is the same as for version A
105	H	At least one subject with baseline poly at position 105	
106	I	No subjects with baseline poly at position 106	Table not produced as the result is the same as for version A
109	J	At least one subject with baseline poly at position 109	
110	K	No subjects with baseline poly at position 110	Table not produced as the result is the same as for version A
118	L	No subjects with baseline poly at position 118	Table not produced as the result is the same as for version A
124	M	No subjects with baseline poly at position 124	Table not produced as the result is the same as for version A
127	N	No subjects with baseline poly at position 127	Table not produced as the result is the same as for version A
128	O	No subjects with baseline poly at position 128	Table not produced as the result is the same as for version A

Treatment Response Tables: Details

Input specifications

Select Population

Intent-to-Treat

Select subgroup variable

Genotype

Label for Subgroup on output

Genotype

Select Subgroup Level

All

Select Level

AMINO ACID

Select Analysis Type

Position (details with mixture notation)

Select Response Variables (if nothing selected, all will be presented)

Nothing selected

Select Gene region

R1

Summary

Details

Select Single position



23

Show table for selected position

TEFPMAAUBLPS103A-ITT: Subgroup Analysis for Baseline Polymorphism at 15 Region 1 Positions of Interest Amino Acid by Subgroups (Genotype) Position 23; Intent-to-Treat (Study PhUSE 2022 Demo)

	SubGroup	Response Parameter	Treatment 1	Treatment 2
Analysis set: Intent-to-Treat	Analysis set: Intent-to-Treat		45	48
Genotype : All Subjects	Genotype : All Subjects		45	48
Subjects with Sequencing Data (R1)	Genotype : All Subjects		15	17
Virus DNA < LLOQ for target detected and not detected at Week 24	Genotype : All Subjects	Virus DNA < LLOQ for target detected and not detected at Week 24		
no baseline Polymorphism at position 23	Genotype : All Subjects	Virus DNA < LLOQ for target detected and not detected at Week 24	8/15 (40.0%)	7/15 (46.7%)
Virus DNA < LLOQ for target detected and not detected at Week 48	Genotype : All Subjects	Virus DNA < LLOQ for target detected and not detected at Week 48		
no baseline Polymorphism at position 23	Genotype : All Subjects	Virus DNA < LLOQ for target detected and not detected at Week 48	12/15 (80.0%)	13/15 (86.7%)
Virus DNA >= 1 log decreases from baseline at Week 4	Genotype : All Subjects	Virus DNA >= 1 log decreases from baseline at Week 4		
no baseline Polymorphism at position 23	Genotype : All Subjects	Virus DNA >= 1 log decreases from baseline at Week 4	13/13 (100.0%)	14/15 (93.3%)
Virus DNA >= 3 log decreases from baseline at Week 24	Genotype : All Subjects	Virus DNA >= 3 log decreases from baseline at Week 24		
no baseline Polymorphism at position 23	Genotype : All Subjects	Virus DNA >= 3 log decreases from baseline at Week 24	12/15 (80.0%)	13/15 (86.7%)
Virus DNA >= 3 log decreases from baseline at Week 48	Genotype : All Subjects	Virus DNA >= 3 log decreases from baseline at Week 48		
no baseline Polymorphism at position 23	Genotype : All Subjects	Virus DNA >= 3 log decreases from baseline at Week 48	12/15 (80.0%)	14/15 (93.3%)
NA Treatment Completion Criteria Met at Week 48	Genotype : All Subjects	NA Treatment Completion Criteria Met at Week 48		
no baseline Polymorphism at position 23	Genotype : All Subjects	NA Treatment Completion Criteria Met at Week 48	1/15 (6.7%)	0/15 (0.0%)

Select list of positions of interest

Graphs

PhUSE 2022 NGS

Tables

>> Single Region

Exploration

Combined Viral Load and PF exploration

Apply Inputs

Input Specs 1

Select Reference Sequence

Universal

VL Profiles and PF DataFilter

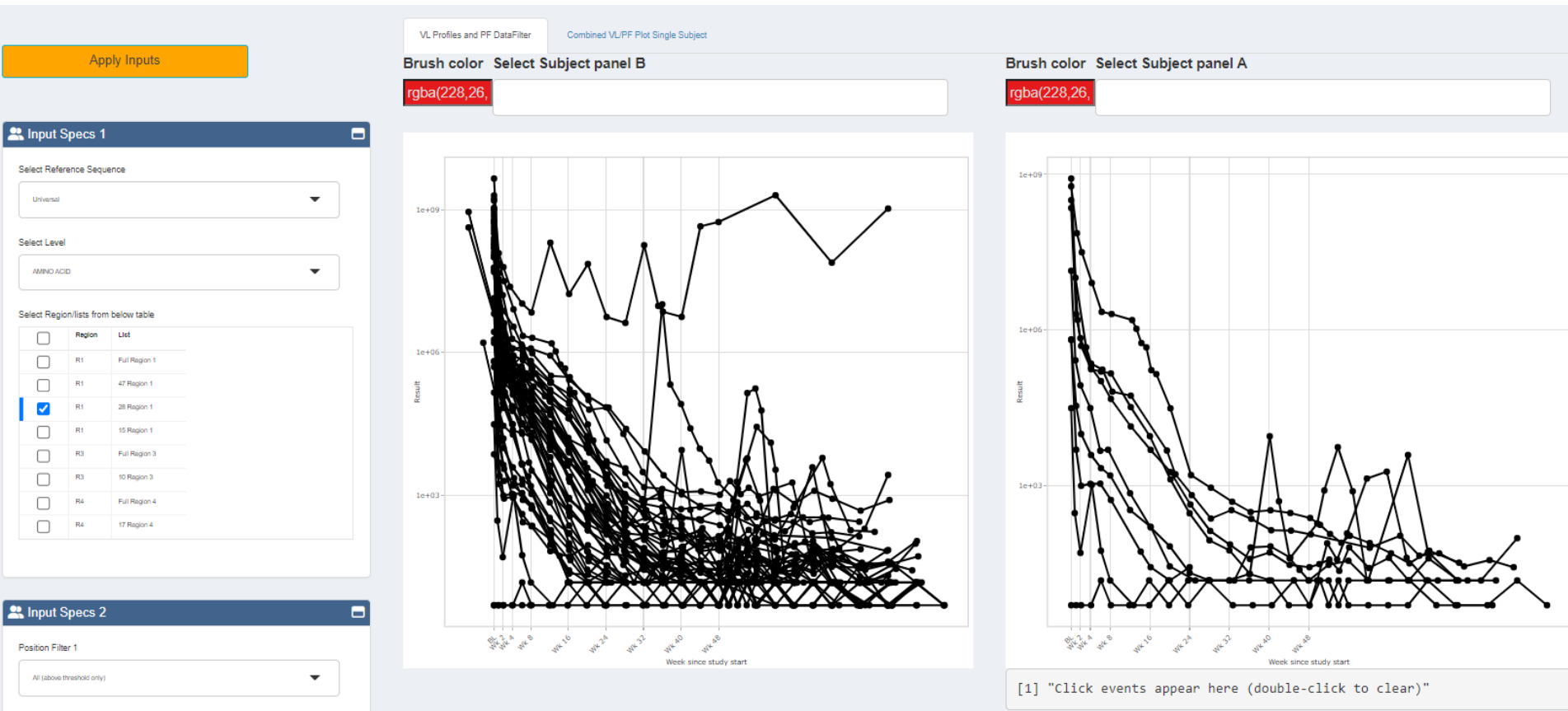
Combined VL/PF Plot Single Subject

Brush color

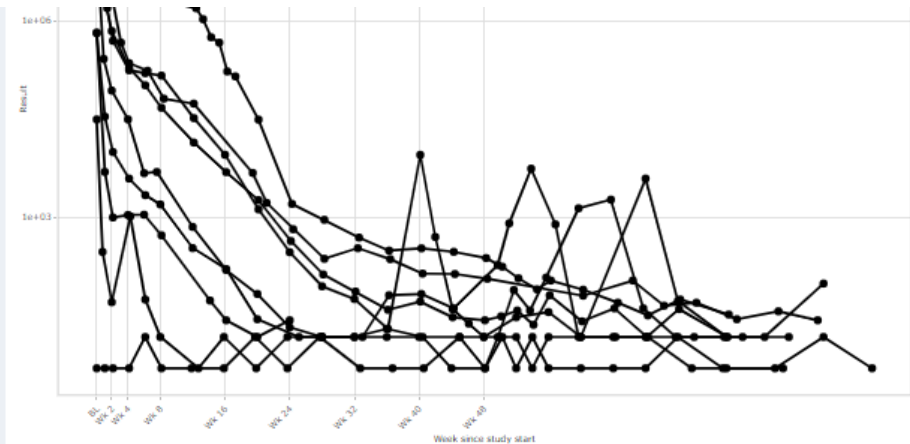
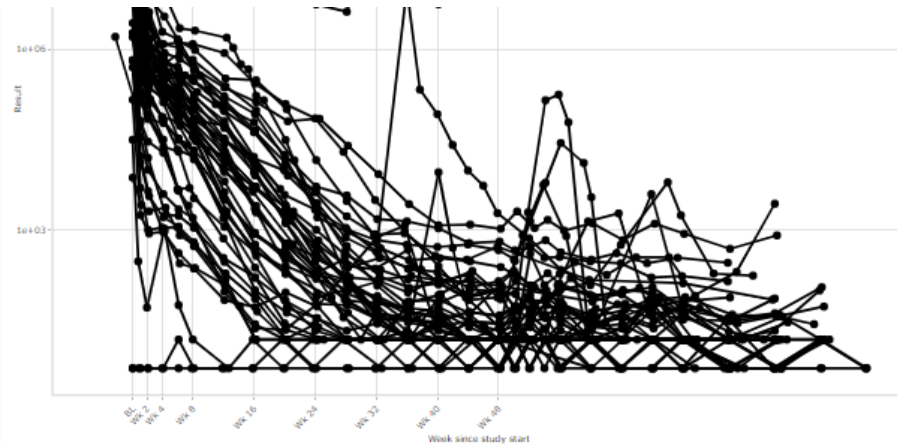
rgba(228,26,228,1)

Select Subject panel B

Graphs



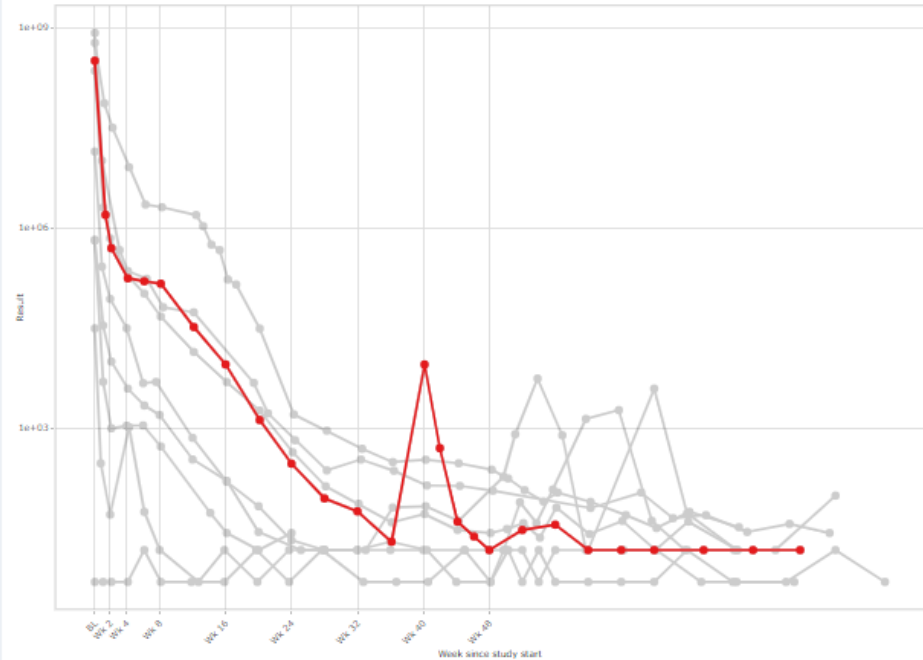
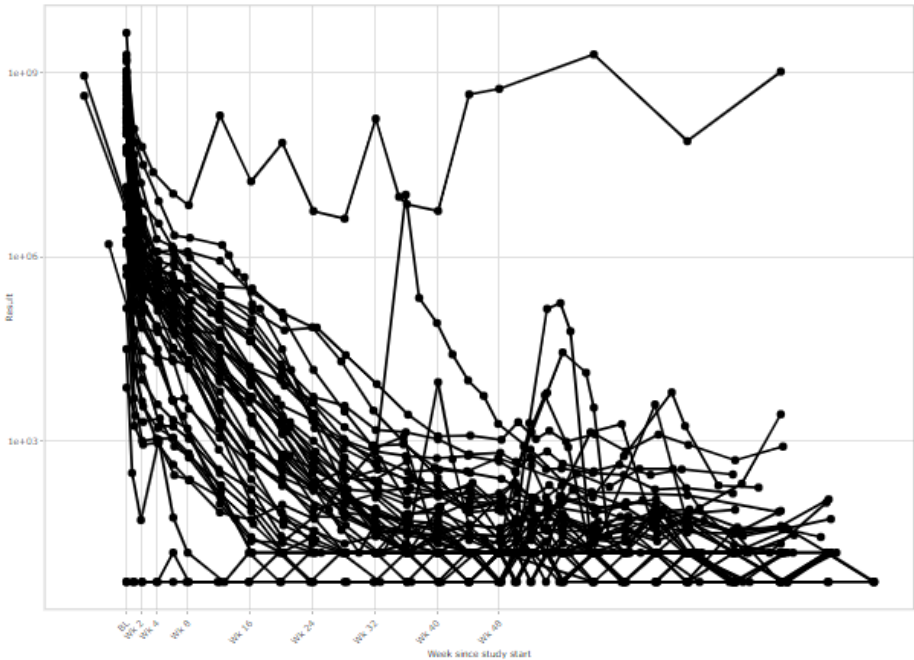
Graphs



[1] "Click events appear here (double-click to clear)"

Subject	Parameter	Region	Position	Tp1	Tp2	Tp3	Tp4	Tp5
PHUSE2022-00027				BL	WV6			
	DNA			15800000	4840			
	SAG			16032.31	231.85			
		R1	000					
PHUSE2022-00160			133-R	R/K	KJR			
				BL	WV12			
	DNA			1.87e+06	54500			
	SAG			1371+17.5	6167.33			
PHUSE2022-00188			000					
			38-Y	YH	YH			
				BL	WV6	WV10		
	DNA			1.87e+06	146000	9030		

Zoom in on selected subject



[1] "Selected subject is : PHUSE2022-100189"

Subject	Parameter	Region	Position	Tp1	Tp2	Tp3
PHUSE2022-100189				BL	Wk8	Wk40
	DNA			1.87e+08	146000	9020
	SAG			60752.07	4880.44	203.06
		R1	000	-	-	
			102-00			WSTOP

The technical side

Transmission electron micrograph of AIDS_HIV-1

Shiny

- Open source R package
- Build Web Applications using R
- Interactive
- Easy to use by End User



How we went from static to dynamic tables

- Server side: Separate app code from tool/content code - Put all the static code for table creation in a function.
- UI side: restrict inputs to VALID combinations only
- UI side: Add **action buttons** to prevent immediate responsiveness when input fields are populated

UI side: restrict inputs to **VALID** combinations only

Eg : Region1 3 lists, region 3 only 1, region 2 none, ..., input options for list depend on selected region

All options

level	region	PoslistShc	PoslistFull	PoslistNum
AMINO A	Region 1	S0	Full Region 1	
AMINO A	Region 1	S1	47 Region 1	18, 19, 23, 24,
AMINO A	Region 1	S2	28 Region 1	23, 24, 25, 29,
AMINO A	Region 1	S3	15 Region 1	23, 24,
AMINO A	Region 2	S0		
AMINO A	Region 2	S1		
AMINO A	Region 2	S2		
AMINO A	Region 2	S3		
AMINO A	Region 3	S0	Full Region 3	
AMINO A	Region 3	S1	10 Region 3	169, 173, ...
AMINO A	Region 3	S2		
AMINO A	Region 3	S3		
AMINO A	Region 4	S0	Full Region 4	
AMINO A	Region 4	S1	17 Region 4	116, 118,
AMINO A	Region 4	S2		
AMINO A	Region 4	S3		
AMINO A	Region 5	S0		

Invalid Options

Valid Options Only

level	region	PoslistShc	PoslistFull	PoslistNum
AMINO ACID	Region 1	S0	Full Region 1	
AMINO ACID	Region 1	S1	47 Region 1	18, 19, 23, 24,
AMINO ACID	Region 1	S2	28 Region 1	23, 24, 25, 29,
AMINO ACID	Region 1	S3	15 Region 1	23, 24,
AMINO ACID	Region 3	S0	Full Region 3	
AMINO ACID	Region 3	S1	10 Region 3	169, 173, ...
AMINO ACID	Region 4	S0	Full Region 4	
AMINO ACID	Region 4	S1	17 Region 4	116, 118,
AMINO ACID	Region 9	S0	Full Region 9	
NUCLEOTIDE	Region 1	S0	Full Region 1	
NUCLEOTIDE	Region 2	S0	Full Region 2	
NUCLEOTIDE	Region 3	S0	Full Region 3	
NUCLEOTIDE	Region 4	S0	Full Region 4	
NUCLEOTIDE	Region 4	S1	20 Region 4	261, 262, ...
NUCLEOTIDE	Region 5	S0	Full Region 5	
NUCLEOTIDE	Region 5	S2	20 Region 5	1781, 1782, ...
NUCLEOTIDE	Region 6	S0	Full Region 6	

SelectInput/updateSelectInput or uiOutput

Separate code logic from data logic :

- data logic: define lookup table for valid options
- code logic : utilize lookup table for input selection

Reactable

- take advantage of **R tools html/webpage interactivity** when presenting tables
- Interactive data tables for R -- reactable
- basic interactivity only : sort, filter, pagesize, show page info, ...
- Built-in column formatting

	Treatment 1	Treatment 2	Treatment 3	Treatment 4	Treatment 5
Analysis set: Intent-to-Treat	45	48	93	93	96
Genotype : All Subjects	45	48	93	93	96
Subjects with Sequencing Data (R1)	15	17	32	32	36
no baseline polymorphism at positions of interest	12/15 (80.0%)	12/17 (70.6%)	25/32 (78.1%)	29/32 (90.6%)	31/36 (86.1%)
1 or More Baseline Polymorphism at positions of interest	3/15 (20.0%)	5/17 (29.4%)	7/32 (21.9%)	3/32 (9.4%)	5/36 (13.9%)
24	1/15 (6.7%)	1/17 (5.9%)	4/32 (12.5%)	0	0
25	0	0	1/32 (3.1%)	0	0
29	1/15 (6.7%)	1/17 (5.9%)	0	0	1/36 (2.8%)
30	0	1/17 (5.9%)	0	0	0
33	1/15 (6.7%)	0	0	0	0
105	1/15 (6.7%)	3/17 (17.6%)	4/32 (12.5%)	2/32 (6.2%)	4/36 (11.1%)
109	0	1/17 (5.9%)	0	1/32 (3.1%)	2/36 (5.6%)
Genotype : Genotype A	5	2	9	7	8
Subjects with Sequencing Data (R1)	2	1	6	5	7
no baseline polymorphism at positions of interest	2/2 (100.0%)	0	5/6 (83.3%)	4/5 (80.0%)	5/7 (71.4%)
1 or More Baseline Polymorphism at positions of interest	0	1/1 (100.0%)	1/6 (16.7%)	1/5 (20.0%)	2/7 (28.6%)
24	0	0	1/6 (16.7%)	0	0
105	0	1/1 (100.0%)	1/6 (16.7%)	1/5 (20.0%)	1/7 (14.3%)

Efficacy Response by baseline polymorphism

TEFPMAAUBLPS103-ITT: Subgroup Analysis for Baseline Polymorphism at 15 Region 1 Positions of Interest Amino Acid by Subgroups (Genotype); Intent-to-Treat (Study PhUSE 2022 Demo)

Summary	A (pos 23)	B (pos 24)	C (pos 25)	D (pos 29)	E (pos 30)	F (pos 33)	H (pos 105)	J (pos 109)
TEFPMAAUBLPS103-ITT: Overview table								
POSITION	POSITIONL	POSITIONInfo	note					
23	A	No subjects with baseline poly at position 23						
24	B	At least one subject with baseline poly at position 24						
25	C	At least one subject with baseline poly at position 25						
29	D	At least one subject with baseline poly at position 29						
30	E	At least one subject with baseline poly at position 30						
33	F	At least one subject with baseline poly at position 33						
37	G	No subjects with baseline poly at position 37	Table not produced as the result is the same as for version A					
105	H	At least one subject with baseline poly at position 105						
108	I	No subjects with baseline poly at position 108	Table not produced as the result is the same as for version A					
109	J	At least one subject with baseline poly at position 109						
110	K	No subjects with baseline poly at position 110	Table not produced as the result is the same as for version A					
118	L	No subjects with baseline poly at position 118	Table not produced as the result is the same as for version A					
124	M	No subjects with baseline poly at position 124	Table not produced as the result is the same as for version A					
127	N	No subjects with baseline poly at position 127	Table not produced as the result is the same as for version A					
128	O	No subjects with baseline poly at position 128	Table not produced as the result is the same as for version A					

* Each position with at least one subject with baseline polymorphism will have a separate tab

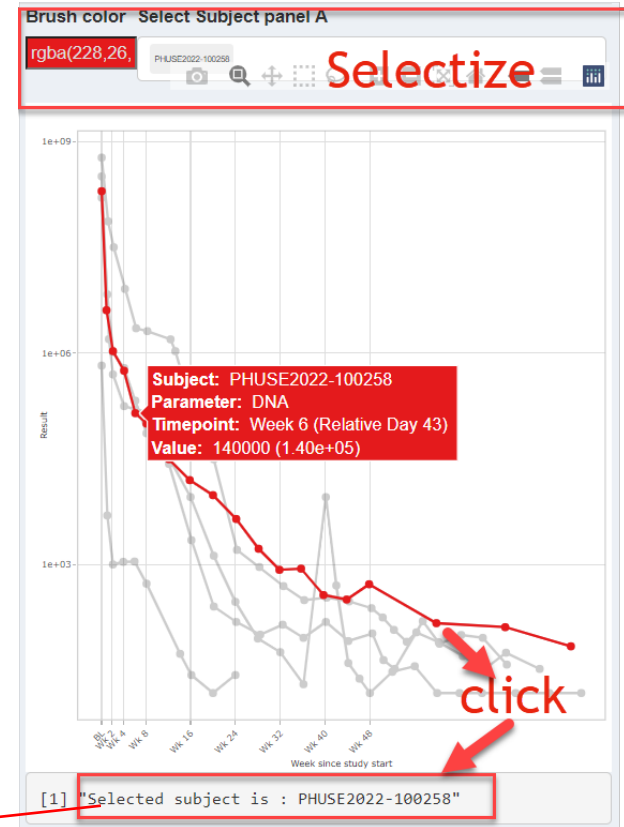
† Only the first position with no subjects with baseline polymorphism will have a separate tab

Download output

- Response by baseline status:
 - Goal : One table per baseline position status
 - Perform derivations for **all positions** of the list at once
 - Download as a **multiple tabs html** file
- Download report : `downloadHandler(filename,content)`
 - Filename : derived identifier
 - Content :
 - **rmarkdown render** process
 - For simpler tables: apply function to save a single rtf/html page
 - <https://mastering-shiny.org/action-transfer.html#downloading-reports>
 - Disable action buttons (download/actions) to prevent multiple launches of time intensive processes (`initiate shinyjs : useShinyjs()`)

Graphs

- Interactive plots :
 - **ggplotly** : tooltip = "text" for customized hover
 - Plotly::highlight , click or selectize
 - Pass selected subject to table :
`plotly::event_data(event='plotly_click')`
- [Plotly.com/r/getting-started](https://plotly.com/r/getting-started)



Subject	Parameter	Region	Position	Tp1	Tp2	Tp3	Tp4
PHUSE2022-100258				BL	Wk12	Wk24	Wk32
	DNA			1.87e+08	31100	4420	840
	SAG			38335.51	27413.29	30332.67	27007.28
		R1	000	-	-		-
			136-N			D1N	

Interactive table – hover info



Phuse_table1.html

PHUSE2022-100316		BL	Wk4	Wk8
DNA		662000	1090	534
SAG		711.18	2565.3	
	R1	000		
		105-I	I T	I T
				I/T

Position = 105 Reference = I
I (76.7%),
T^{\$} (23.3%)

- Hover info in tables :
 - dataframe with main text and hover text
 - reactable **cell render** function with as core **tippy::tippy**
 - reactable **colDef** for single column to apply this cell render function
 - **lapply** all columns

What we learned

- Challenge to have all inputIDs and outputIDs unique across the entire app
- More complicated apps are better organized by using modules
- Separate content/tool code from App code
- Code organization
- Development starts simple, but soon gets more complicated and messy
- Stackoverflow is our best friend!

Conclusion

- Without previous experience with Shiny, we were able to create a useful app, that covered the needs of the scientists
- Study programmers had only minimal preparation work for this deep sequencing part -> focus on other work
- The App was easily transferred from one trial to another, even with some important design element differences
- Applying to another virus should require mainly adjustment in data setup, less in code logic updates

Questions



Contact information:

Dr. Annemie Van Dijck, PhD, MSc

Clinical & Statistical Programming (C&SP)



avandijc@its.inj.com

Dr. Ilse Augustyns, PhD, MSc

Methodology and Innovation (C&SP)



iaugusty@its.inj.com

Integrated Data Analytics & Reporting (IDAR)

Janssen R&D



PHARMACEUTICAL COMPANIES OF *Johnson & Johnson*