PASeq.org: One-click, Cloud-based Web Service for NGS-based HIV genotyping Data Analysis

Marc Noguera-Julian, PhD

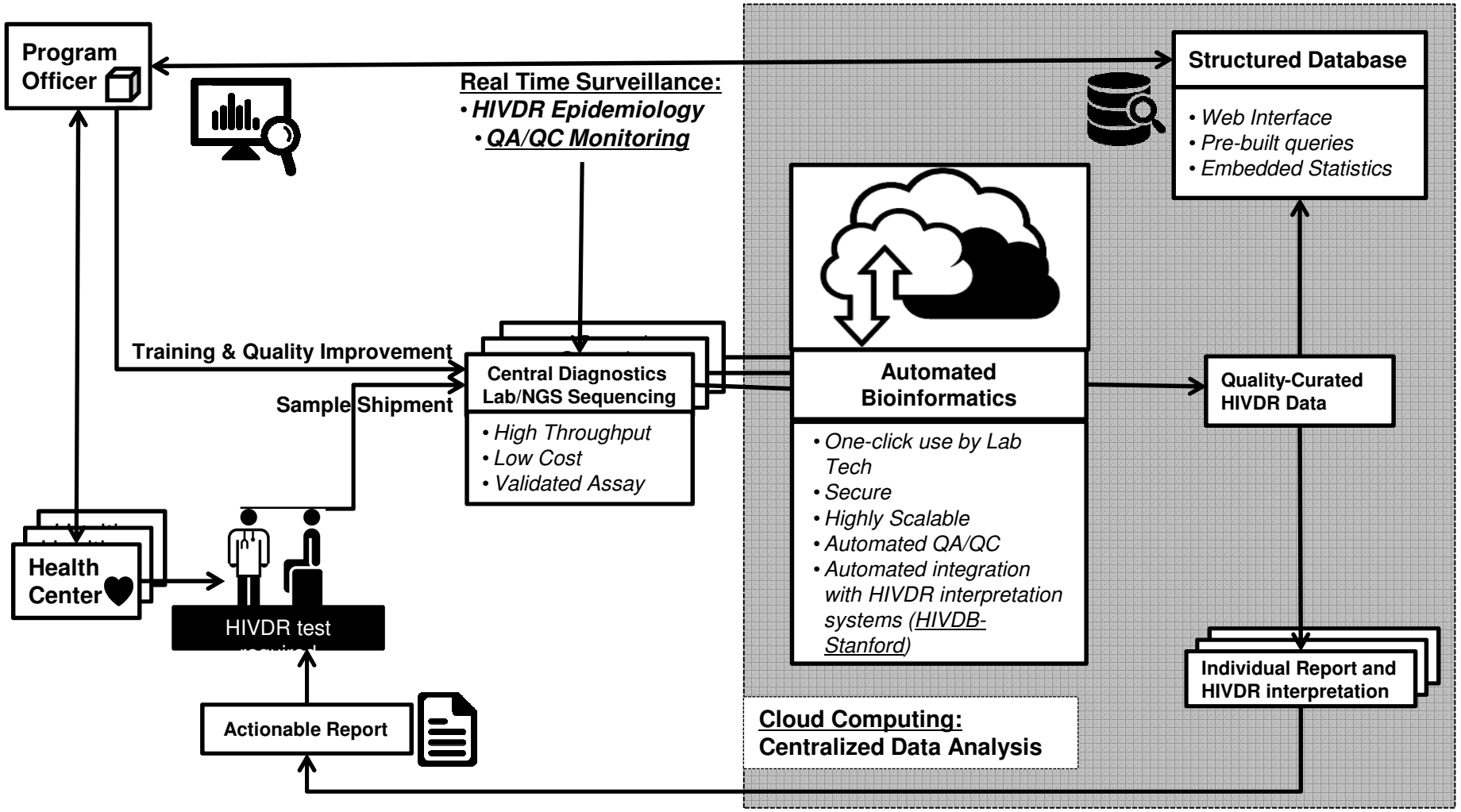
IrsiCaixa AIDS Research Institute

Badalona, Catalonia

IrsiCaixa
Institut de Recerca de la Sida

NGS Data Deluge

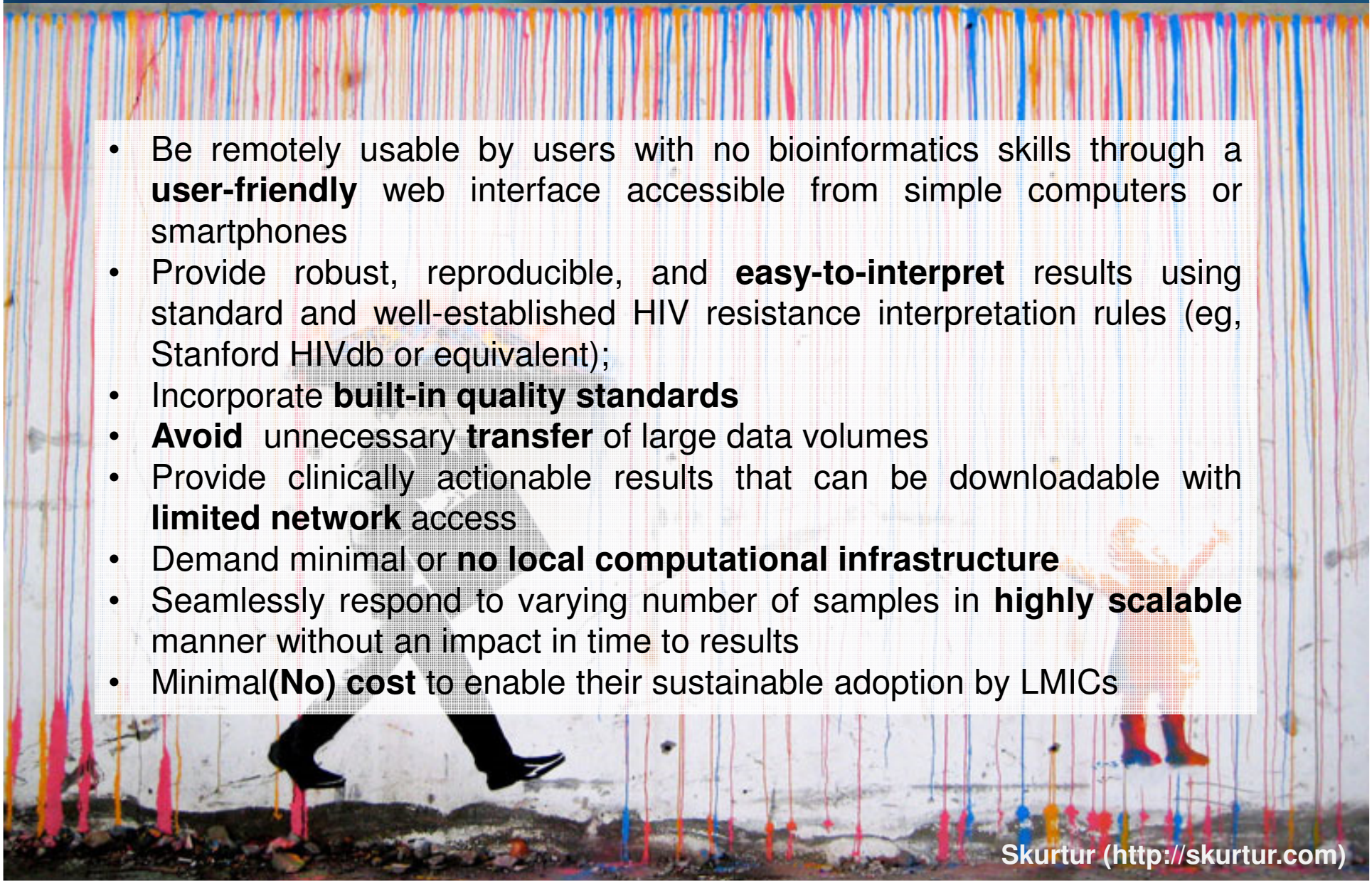
NGS in HIV DR Testing scale up. How?



Noguera-Julian. et al. J. Infect. Dis. 2017

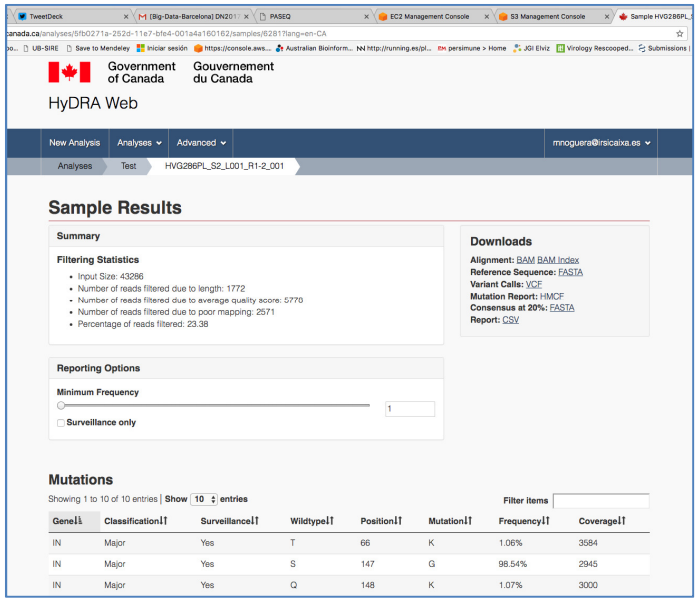
NGS Data Deluge
HIV-NGS-DR Testing Needs

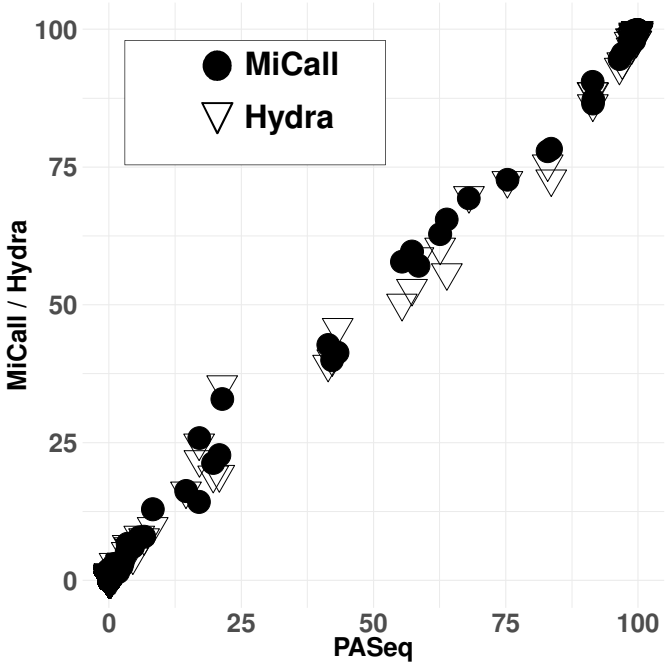
- Be remotely usable by users with no bioinformatics skills through a **user-friendly** web interface accessible from simple computers or smartphones
- Provide robust, reproducible, and **easy-to-interpret** results using standard and well-established HIV resistance interpretation rules (eg, Stanford HIVdb or equivalent);
- Incorporate **built-in quality standards**
- **Avoid** unnecessary **transfer** of large data volumes
- Provide clinically actionable results that can be downloadable with **limited network** access
- Demand minimal or **no local computational infrastructure**
- Seamlessly respond to varying number of samples in **highly scalable** manner without an impact in time to results
- Minimal**(No) cost** to enable their sustainable adoption by LMICs



Available software for HIV-NGS-DR Testing

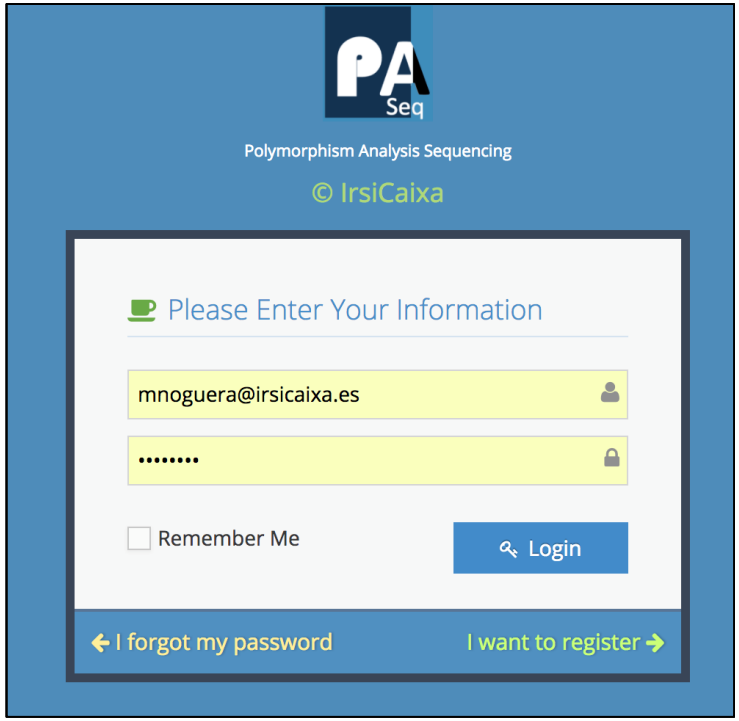
Software	URL	Cost / Sample	Time	Bioinformatic - IT needs	Cloud Based	Year	Commercial use	Web Interface
V-Phaser[19]	https://www.broadinstitute.org/viral-genomics/v-phaser-2	Free		Yes	No	2013	No	No
ShoRah[26]	https://github.com/cbg-ethz/shorah	Free		Yes	No	2013	No	No
VirVarSeq[27]	https://sourceforge.net/projects/virtools/	Free		Yes	No	2015	No	No
VirFlow[28]	https://github.com/ICONIC-UCL/vir-flow	Free	3h	Yes	Yes[c]	2016	No	No
MinVar[a][29]	https://ozagordi.github.io/MinVar/	Free	<1h	Yes	Yes[c]	2016	No	No
MiCall[a]	https://github.com/cfe-lab/MiCall	Free		No	Yes	2016	No	Yes
HyDRA	https://hydra.canada.ca/	Free	<1h	No	Yes	2016	No	Yes
PASeq[a]	https://www.paseq.org	Free	<1h	No	Yes	2016	No	Yes
SmartGeneHIV	http://www.smartgene.com/mod_ngs.html	N/A		No	No	2016	Yes	No
DeepChekHIV[30]	https://www.ablsa.com/overview/deepchek/	$65	<1h	No	Yes	2014	Yes	Yes





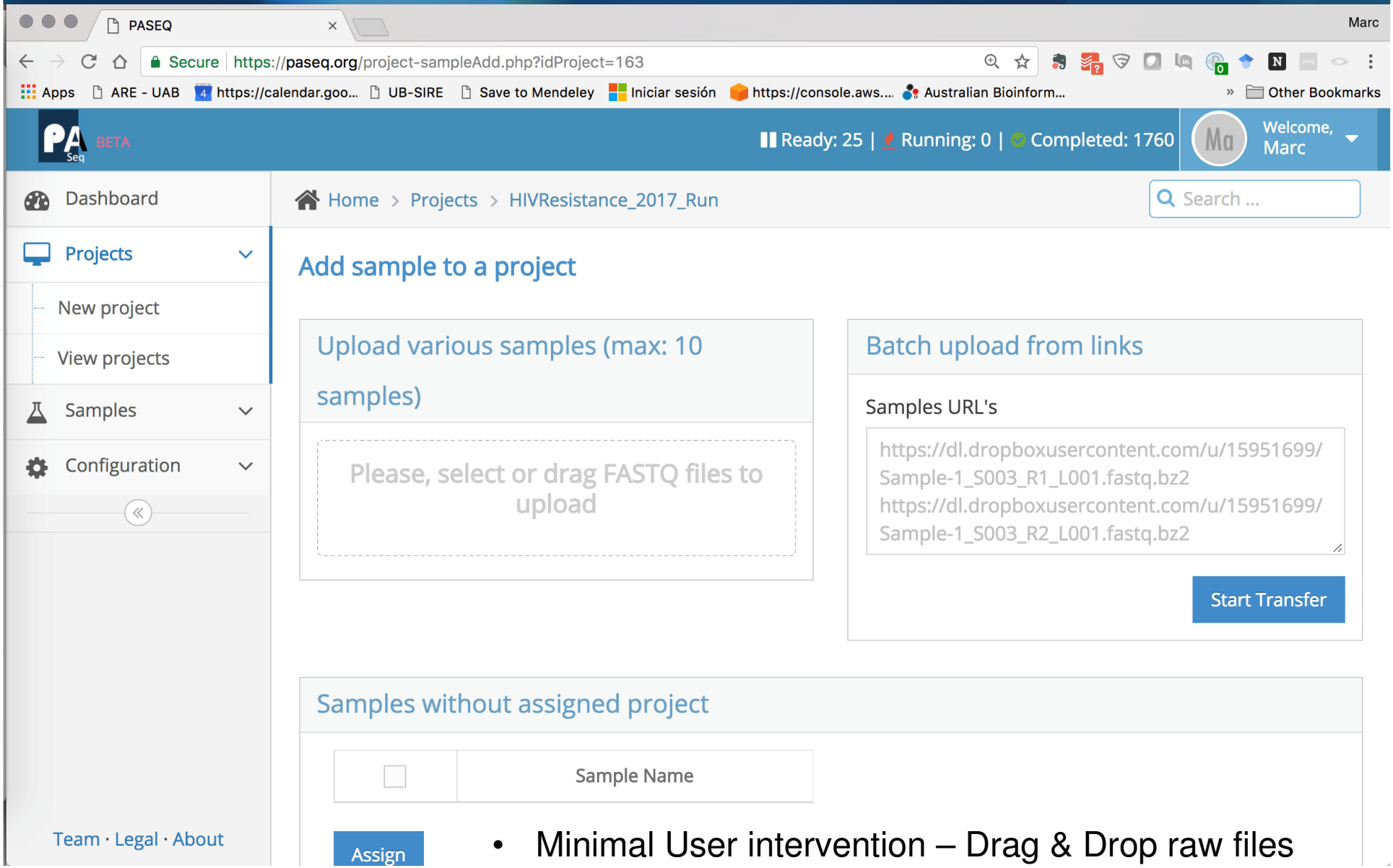
Noguera-Julian. et al. J. Infect. Dis. 2017

PASeq:Polymorphism Analysis by Sequencing



www.paseq.org

HIV-NGS-DR Testing in PASeq: User Friendly Interface



- Minimal User intervention – Drag & Drop raw files
- User can input sample metadata (optional)

HIV-NGS-DR Testing – PASeq Output – Job Logs and duration

2017-11-06 07:27:25. Sample created
2017-11-06 07:28:10. Allocating Resources
2017-11-06 07:28:47. Server assigned. Launching instance
2017-11-06 07:29:40. Initiating process
2017-11-06 07:29:40. Downloading Fastq R1 file
2017-11-06 07:29:40. Downloading Fastq R2 file
2017-11-06 07:29:40. This is a Paired-end Analysis
2017-11-06 07:29:40. Creating WorkSpace
2017-11-06 07:29:41. Going Through Quality Control Using Trimmomatic
2017-11-06 07:31:01. Quality Filtering and Adapter Trimming: Trimmomatic
2017-11-06 07:31:01. MinLen=75 minQual=30 SlidingWindow=20
2017-11-06 07:32:53. 17937 of 21643 survived
2017-11-06 07:32:54. Checking for External Contamination
2017-11-06 07:34:32. Found 17016 HIV sequences and 921 non-HIV sequences
2017-11-06 07:34:40. Indentifying potential contamination source
[illegible] Compressing Files and moving on
2017-11-06 07:36:00. Creating Coverage plots
2017-11-06 07:36:53. Calling Deep Variants
2017-11-06 07:39:59. Querying HIVDB using deep Variant Data at different thresholds
2017-11-06 07:40:49. Querying HIVDB-Stanford with consensus sequence for resistance
2017-11-06 07:40:50. Storing Consensus sequences for surveillance
2017-11-06 07:41:05. Uploading Results
2017-11-06 07:41:42. Job Finished

Data Download & setup

Quality Control

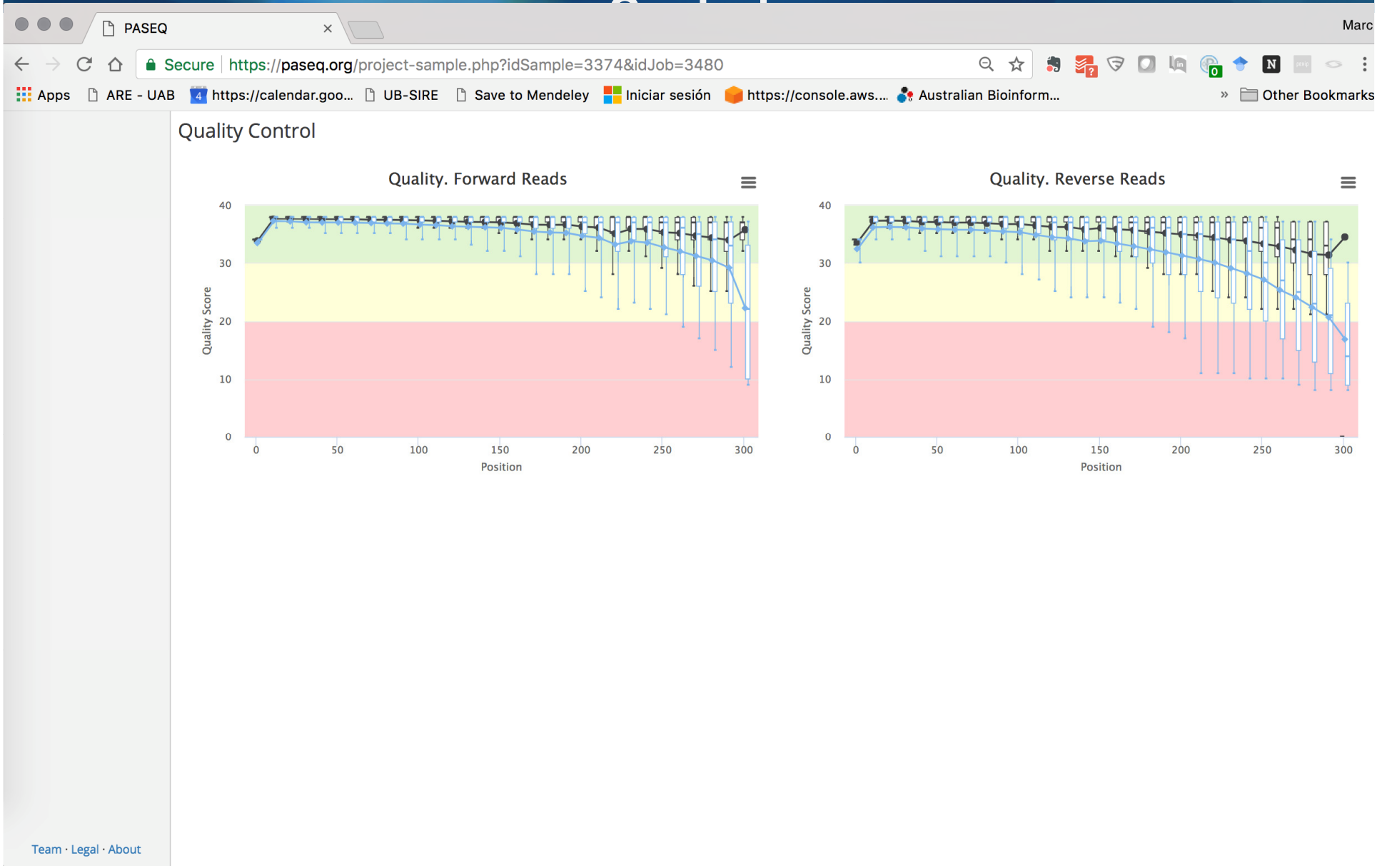
Contamination Control

Sequence Alignment & Variant Calling

Resistance Interpretation

~30 min Analysis Time

HIV-NGS-DR Testing – PASeq Output – Embedded Quality



HIV-NGS-DR Testing – PASeq Output – User-defined threshold

Resistance

Stanford HIVDBVersion:8.2 web service:beta-1.0.1

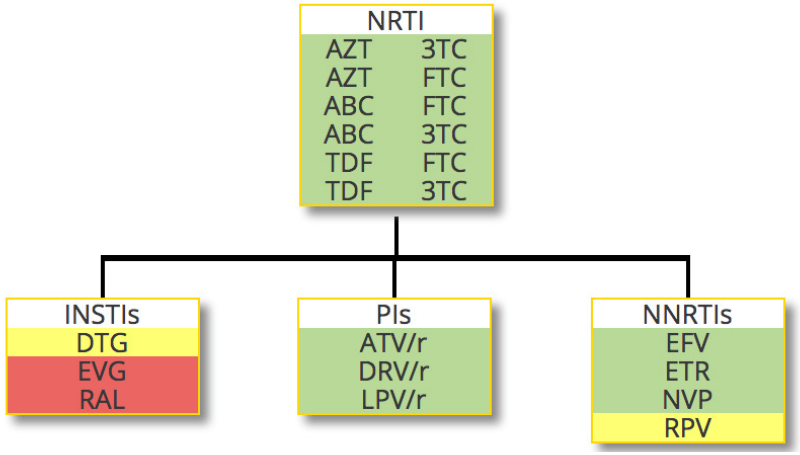
Threshold level: 5%

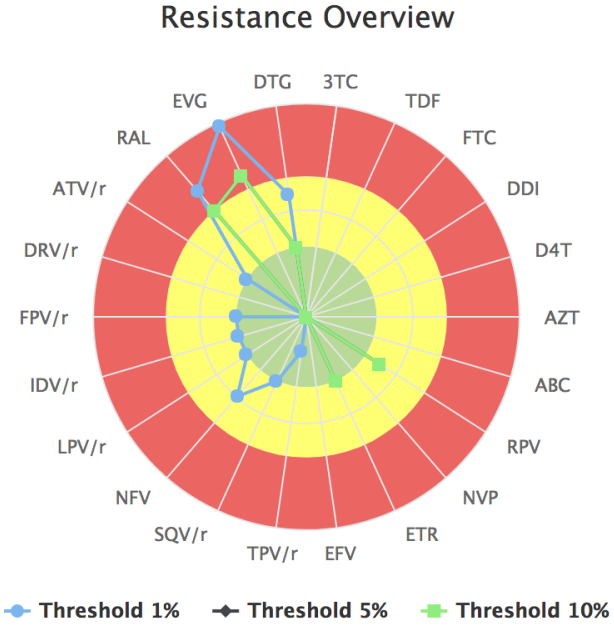
Drug	20%	5%
NRTI		
lamivudine (3TC)		
NNRTI		
efavirenz (EFV)		
etravirine (ETR)	E138A	E138A
nevirapine (NVP)		
rilpivirine (RPV)	E138A	E138A
PI		
atazanavir/r (ATV/r)		
INSTI		
dolutegravir (DTG)	N155H	N155H
elvitegravir (EVG)	N155H,S147G	N155H,S147G
raltegravir (RAL)	N155H	N155H

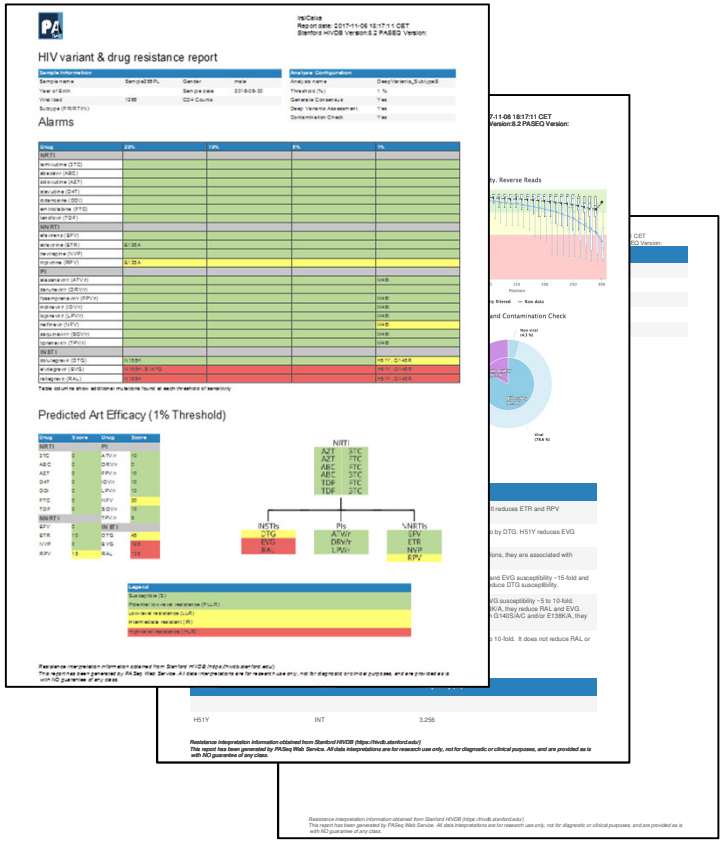
HIV-NGS-DR Testing – PASeq Output – Resistance Interpretation

Treatment Efficacy (Threshold: 1%)

PaSeq version





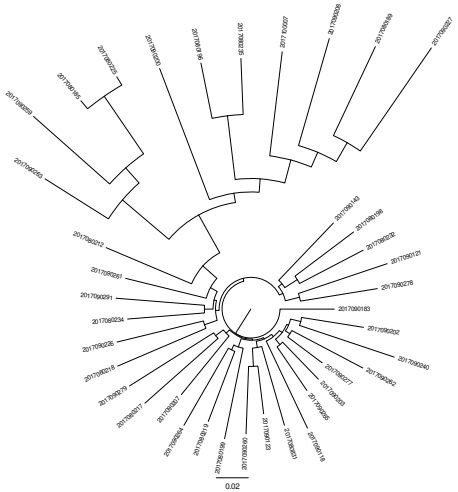


HIV-NGS-DR Testing – PASeq – Data Structure

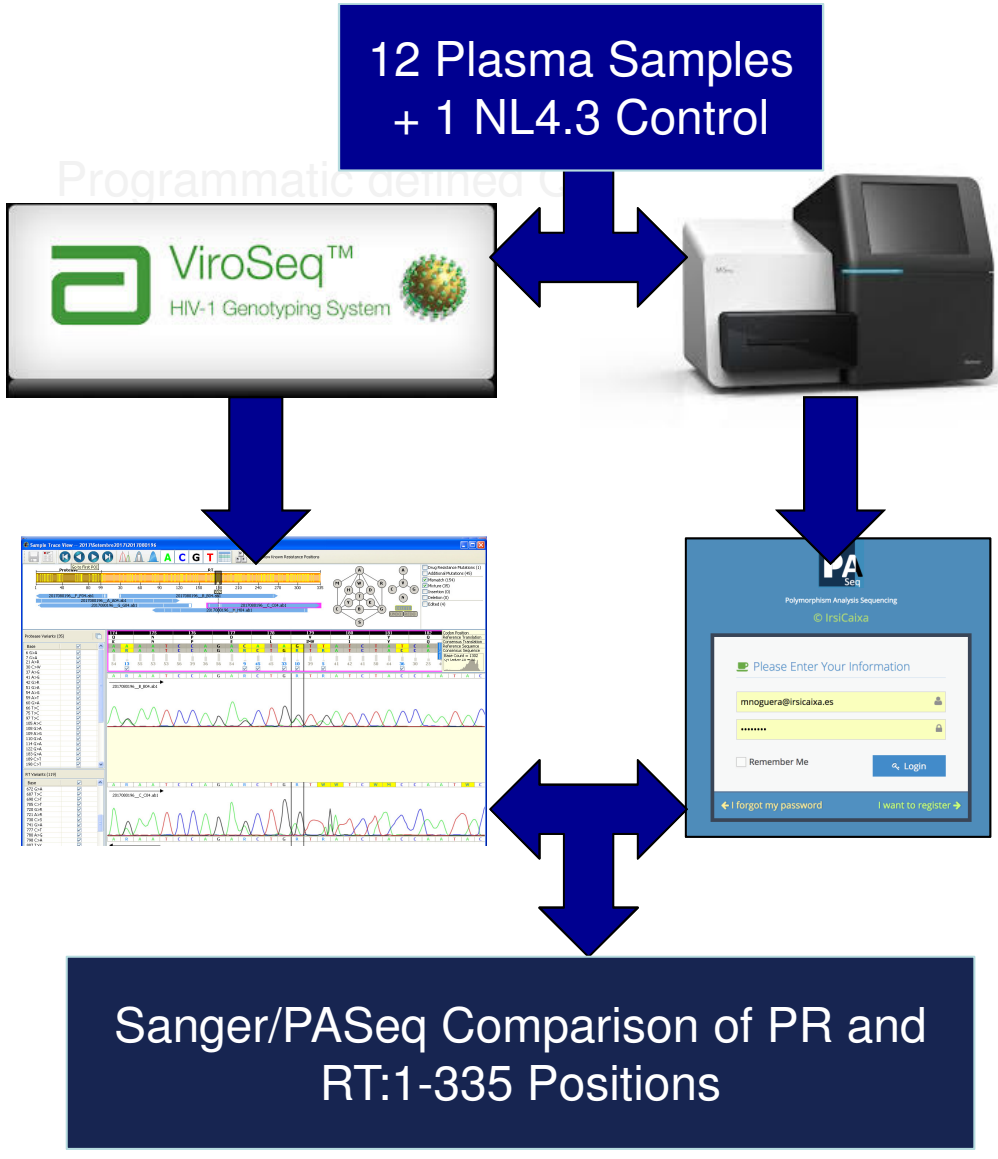
Programmatic defined Queries

Describe **TFV** resistance in samples from project IWDR in **June 2017**

SELECT **ConsensusSequence** from all samples for **project IWDR** that are Intermediate or High resistance to TFV



HIV-NGS-DR Testing – PASeq – Results



All except 5/6555 sanger(0.0001%) aminoacid readouts were detected by NGS/PASeq

Once revised, the 5 discrepant readouts corresponded to low quality or dual mixes sanger results.

PASeq detected 186 additional mutations with a median frequency 3.420%(IQR:1.72-10.69%).

None of the PASeq detected additional mutations was detected in the NL4.3 control.

HIV-NGS-DR Testing – PASeq – Wrap-up

- PASeq is a **Web Service** for HIV resistance analysis using NGS data.
- PASeq is **free to use** and open source.
- Seamless scale-up backed by **cloud computing(AWS)**
- Streamlined pipeline: From raw data to **easy-to-interpret resistance report** with **minimal user interaction**
- Embedded **quality control** and **Stanford HIVDB** queries with user-defined threshold.
- Pre-defined or user defined analysis configuration.
- **Agnostic** to experimental design.
- **Multiplexed** parallel data upload, analysis and results download.
- Database structured results allow for **real-time** data **smart surveillance**.
- **Downloadable PDF** sample report generation

THANKS!

IrsiCaixa
Institut de Recerca de la Sida

Jordi Herms

www.paseq.org

Visit at Poster #20