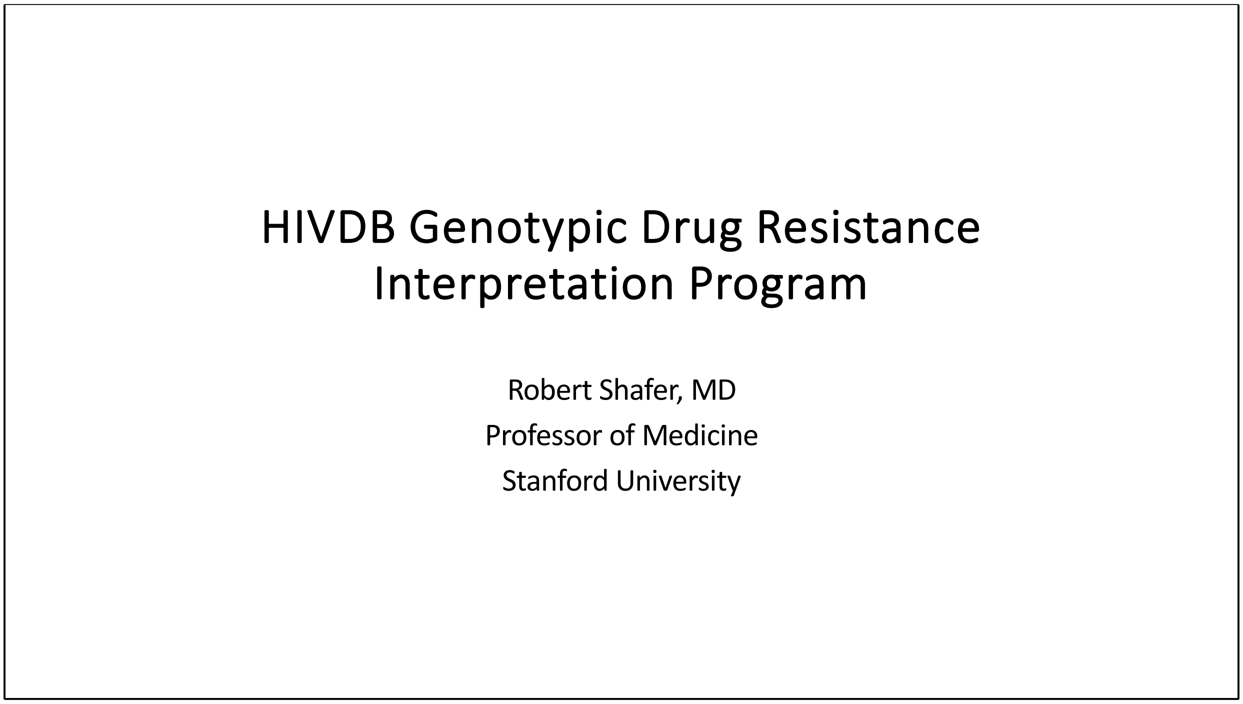
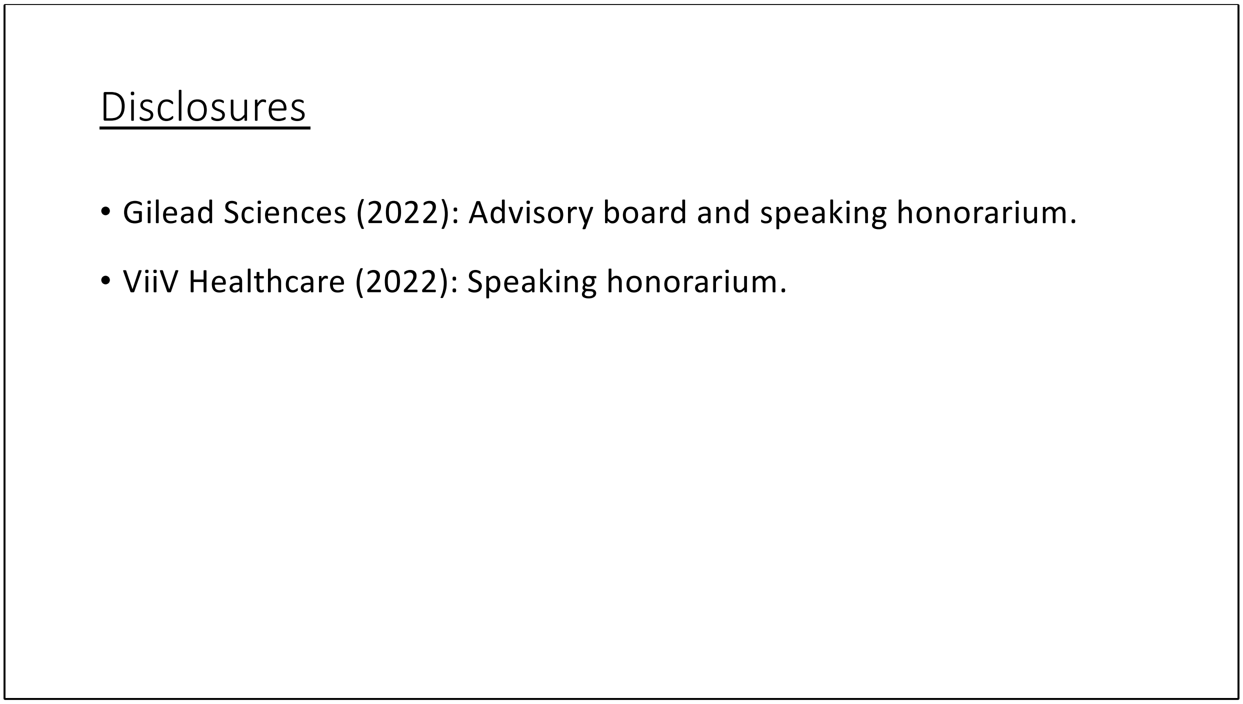


HIVDB Genotypic Drug Resistance Interpretation Program

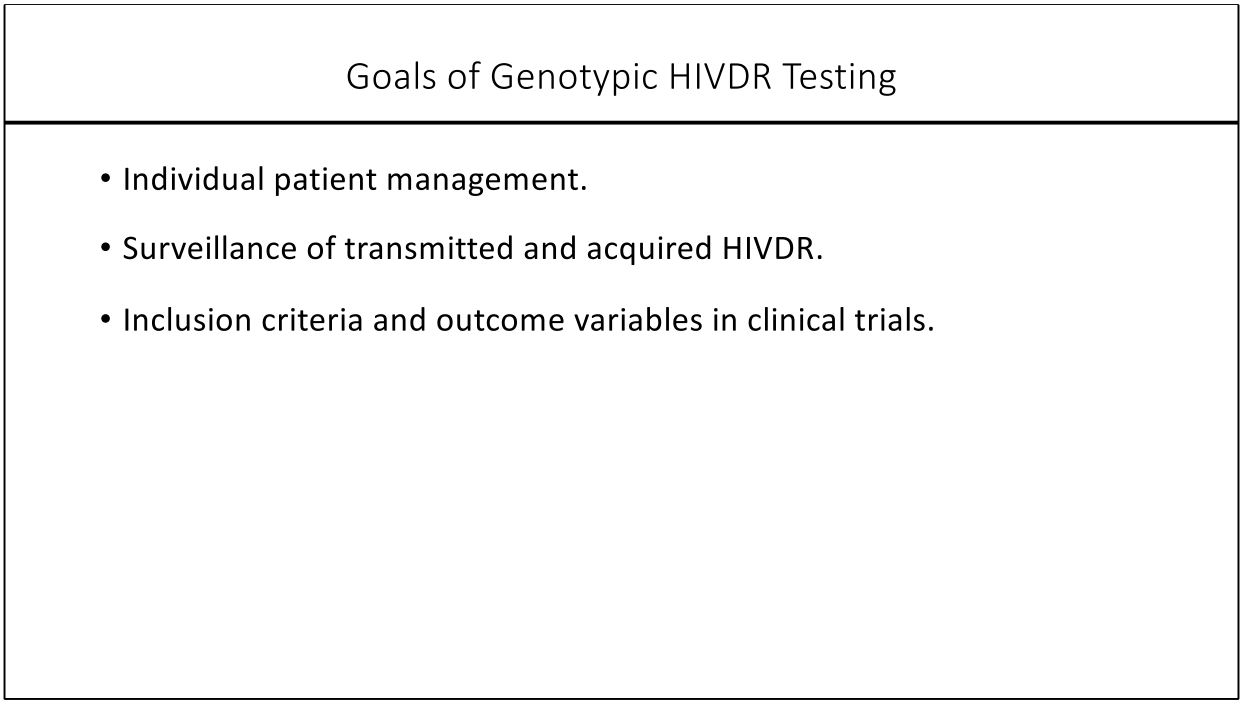
Robert Shafer, MD
Professor of Medicine
Stanford University



Disclosures

- Gilead Sciences (2022): Advisory board and speaking honorarium.
- ViiV Healthcare (2022): Speaking honorarium.

These are my disclosures.

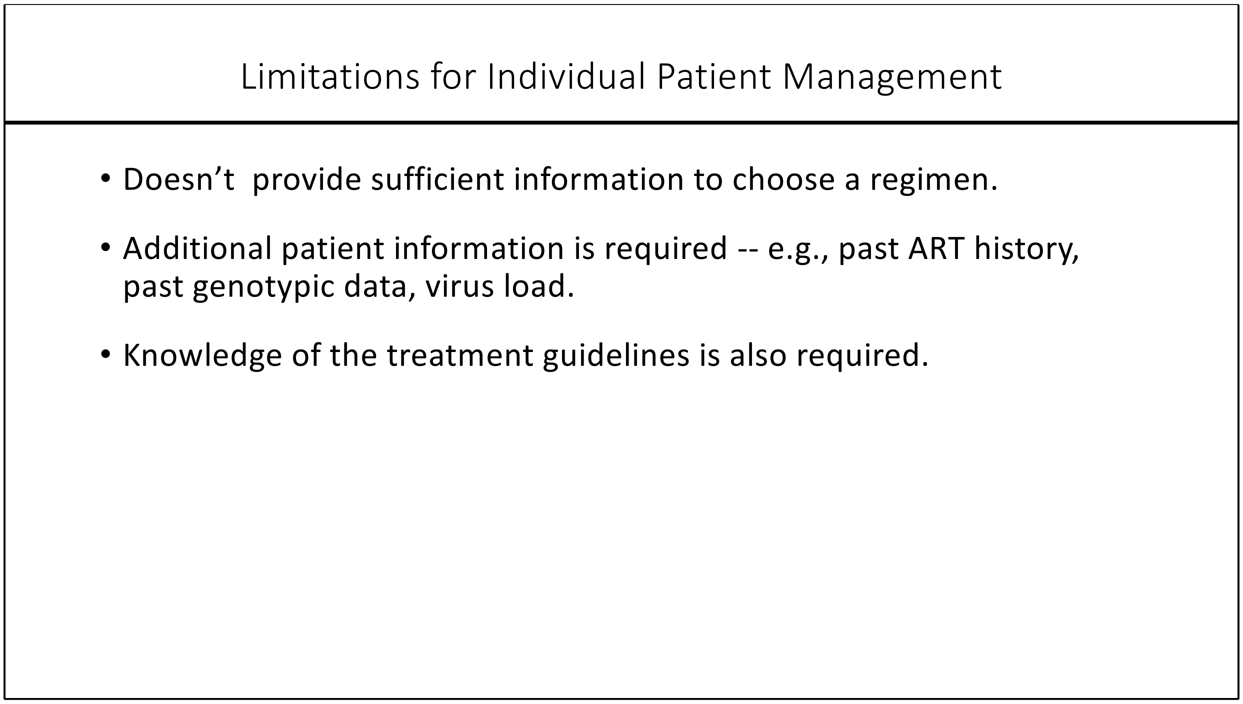


Goals of Genotypic HIVDR Testing

- Individual patient management.
- Surveillance of transmitted and acquired HIVDR.
- Inclusion criteria and outcome variables in clinical trials.

1. Genotypic HIVDR testing is performed for three main purposes.
2. In UICs it is performed to optimize the management of individual patients.
3. In LMICs, it is performed largely for the surveillance of transmitted and acquired HIVDR.
4. In clinical trials, it is used as inclusion criteria and as an outcome variable.
5. For the 2nd and 3rd goals, it has become important to use a consistent approach so that the results of different studies can be compared with one another and can be combined.

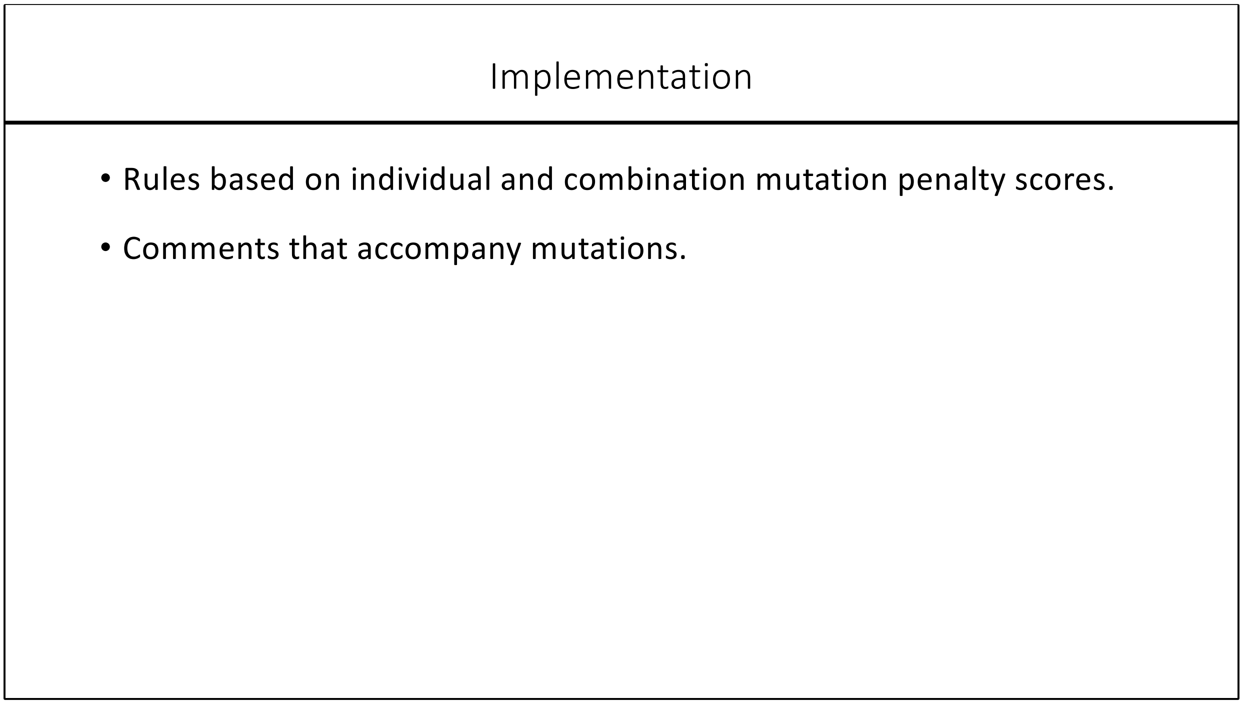
6. The HIVDB interpretation program is frequently used as a standard for the 2nd and 3rd goals.
7. For the 1st goal, the interpretation program is intended to be educational. It has not been approved by the FDA or any other regulatory body.
8. Moreover, the program it lacks the heuristic power to provide specific ARV treatment recommendations because it does not integrate the additional clinical data needed by clinicians to choose therapy.
9. These data include a patient’s past treatment history, previous GRT results, plasma HIV-1 RNA levels, and information on the likelihood of adherence.
10. Finally, the program does not provide sufficient information



Limitations for Individual Patient Management

- Doesn’t provide sufficient information to choose a regimen.
- Additional patient information is required -- e.g., past ART history, past genotypic data, virus load.
- Knowledge of the treatment guidelines is also required.

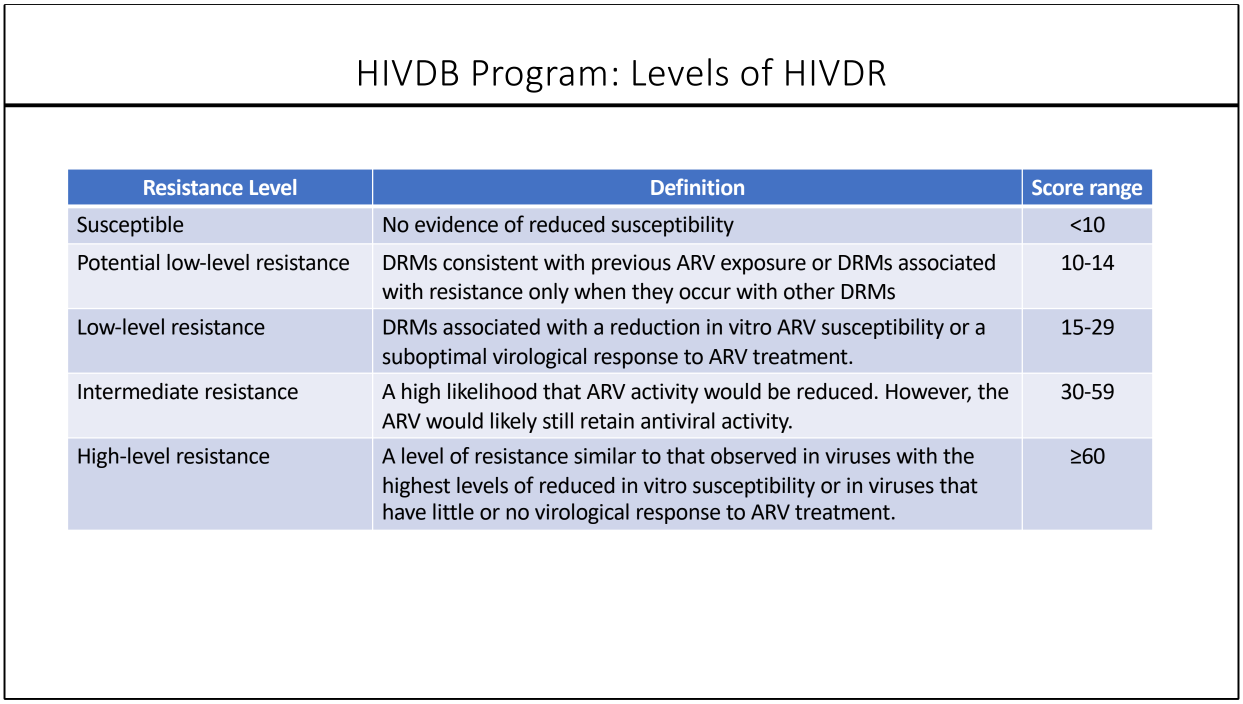
1. For the 1st goal, the interpretation program is intended to be educational. It has not been approved by the FDA or any other regulatory body.
2. Moreover, the program it lacks the heuristic power to provide specific ARV treatment recommendations because it does not integrate the additional clinical data needed by clinicians to choose therapy.
3. These data include a patient’s past treatment history, previous GRT results, plasma HIV-1 RNA levels, and information on the likelihood of adherence.
4. Finally, the program does not provide sufficient information



Implementation

- Rules based on individual and combination mutation penalty scores.
- Comments that accompany mutations.

1. The interpretation system comprises numerical penalties for individual DRMs as well as additional penalties when certain mutations occur in combination.
2. The individual plus combination DRM penalties for a specific ARV drug are added together to derive a total penalty for the drug.
3. Each DRM is accompanied by a comment to provide additional information for how to influences susceptibility to different drugs.
4. Some mutations that were once considered to be DRMs also have comments indicating why we don’t think the DRM is clinically relevant.
5. A rules-based interpretation system seems somewhat archaic in this era of machine learning and AI.
6. However, for reasons that go beyond the scope of this talk, we don’t believe that it is currently possible to develop an optimal ML algorithm for genotypic HIVDR interpretation.



HIVDB Program: Levels of HIVDR

Resistance Level	Definition	Score range
Susceptible	No evidence of reduced susceptibility	<10
Potential low-level resistance	DRMs consistent with previous ARV exposure or DRMs associated with resistance only when they occur with other DRMs	10-14
Low-level resistance	DRMs associated with a reduction in vitro ARV susceptibility or a suboptimal virological response to ARV treatment.	15-29
Intermediate resistance	A high likelihood that ARV activity would be reduced. However, the ARV would likely still retain antiviral activity.	30-59
High-level resistance	A level of resistance similar to that observed in viruses with the highest levels of reduced in vitro susceptibility or in viruses that have little or no virological response to ARV treatment.	≥60

1. Individual DRMs receive penalties that range from 5 to 60.
2. The penalties are titrated so that when the penalties for a genotype are added together, they generally fall in a range from 0 to slightly more than 100.
3. This table indicates how the numerical score is translated into one of 5 different resistance levels from susceptible to high-level resistance.
4. A few DRMs have negative scores for certain drugs. These hyper-susceptibility mutations can mitigate the effect of other DRMs on that drug.
5. The table also indicates what is meant with the different levels of reduced susceptibility.

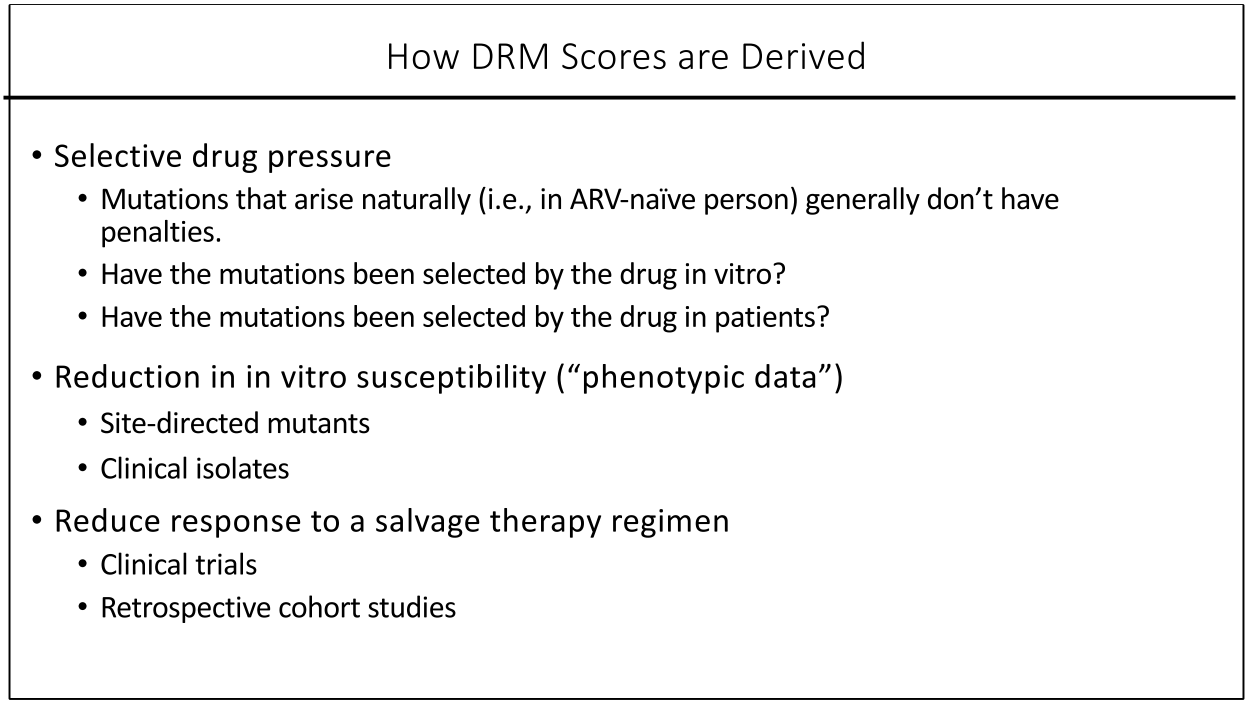
6. "Susceptible" is assigned when a virus displays no evidence reduced susceptibility when compared with a wild-type virus.

7. "Potential low-level resistance" is assigned when a virus has DRMs consistent with previous ARV exposure or contains DRMs associated with resistance only when they occur with other DRMs.

8. "Low-level resistance" is assigned when a virus has DRMs associated with a reduction in vitro ARV susceptibility or a suboptimal virological response to ARV treatment.

9. "Intermediate resistance" is assigned when, although there is a high likelihood that an ARV's activity would be reduced in the presence of a virus's DRMs, the ARV would

likely still retain significant antiviral activity against the virus.
10. "High-level resistance" is assigned when a virus has DRMs predicted to confer a level of resistance similar to that observed in viruses with the highest levels of reduced in vitro susceptibility or in viruses that have little or no virological response to ARV treatment.



How DRM Scores are Derived

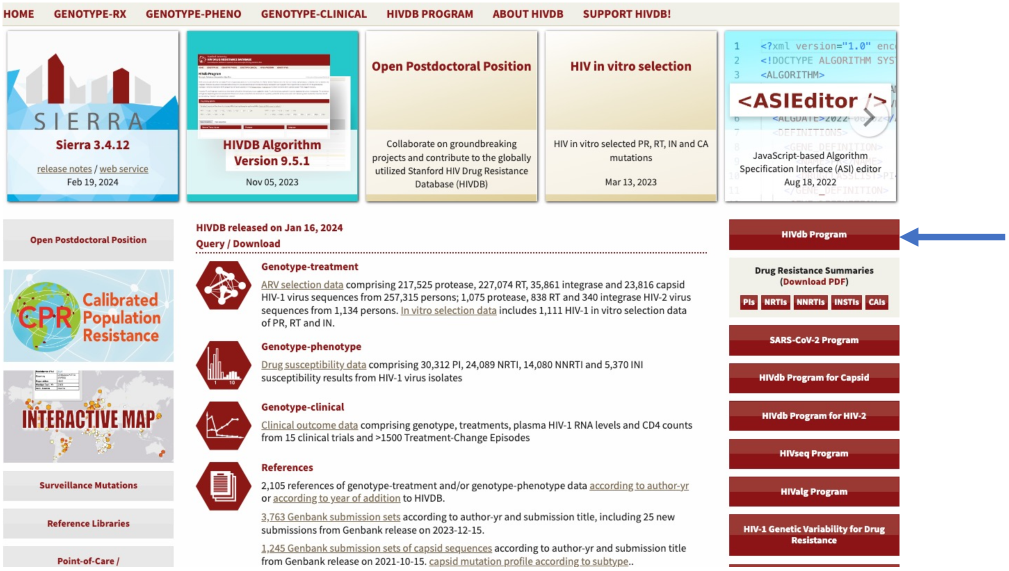
- Selective drug pressure
  - Mutations that arise naturally (i.e., in ARV-naïve person) generally don’t have penalties.
  - Have the mutations been selected by the drug in vitro?
  - Have the mutations been selected by the drug in patients?
- Reduction in in vitro susceptibility (“phenotypic data”)
  - Site-directed mutants
  - Clinical isolates
- Reduce response to a salvage therapy regimen
  - Clinical trials
  - Retrospective cohort studies

1. The DRM penalty scores and comments are created and updated based on based on 3 main types of data/considerations: (i) Are the mutations selected by the drug; (ii) Do the mutations reduce drug susceptibility; (iii) Do the mutations reduce the virological response to a regimen containing the drug; and (iv) what is the consensus expert opinion on how the mutations affects the drug.
2. Now lets drill down into each type of data.
3. First selective drug pressure: Generally, mutations that arise naturally – also referred to as polymorphisms – are generally not assigned penalties or assigned very low penalties.
4. Empirically such mutations have been very rarely associated with reduced drug susceptibility which is fortunate and reflects the fact that for the main classes of drugs resistance does not arise naturally
5. Mutations that are selected in vitro by a drug are often the main mutations that arise in patients, but this is not always the case.
6. Some mutations selected in vitro never occur in patients and frequently mutations that have not been reported in vitro do arise in patients with VF on a drug.
7. Second, in vitro drug susceptibility or phenotypic data: Susceptibility data can be performed on site-directed mutants. In this scenario it is possible to strongly link

the effects of individual mutations or combinations of mutations on reduced drug susceptibility.

8. However, most DRMS and combinations of DRMs have not been studied in this way. Therefore, much of the in vitro susceptibility data comes from clinical isolates. These data are more complex because clinical isolates often contain many different combinations of DRMs and sometimes mutations that are not considered DRMs – often referred to as backbone mutations – can modulate the effect of a DRM combination.
9. Third, virological response to a regimen containing the drug of interest: Such orrelations between genotype and virological suppression are relevant because sustained virological suppression is the main goal of ARV therapy.
10. There have been several highly informative clinical trials that have demonstrated associations between specific pre-therapy DRMs and the risk of VF.
11. However, many retrospective studies have had too few patients relative to the large number of covariates associated with response to therapy – for example, the patient's complete ART history, plasma VL, the drugs used in combination with the drug of interest, and the level of adherence to therapy.
12. Moreover, most retrospective studies have been confounded by the fact that the results of GRT were used to guide the choice of therapy. These data are obtained from the published literature.
13. Finally, for areas in which there is a data vacuum, consensus expert opinion is also used to influence how DRM scores are derived.

HIVDB Home Page => HIVDB Program



Sierra 3.4.12
release notes / web service
Feb 19, 2024

HIVDB Algorithm Version 9.5.1
Nov 05, 2023

Open Postdoctoral Position
Collaborate on groundbreaking projects and contribute to the globally utilized Stanford HIV Drug Resistance Database (HIVDB)

HIV in vitro selection
HIV in vitro selected PR, RT, IN and CA mutations
Mar 13, 2023

JavaScript-based Algorithm Specification Interface (ASI) editor
Aug 18, 2022

HIVDB released on Jan 16, 2024
Query / Download

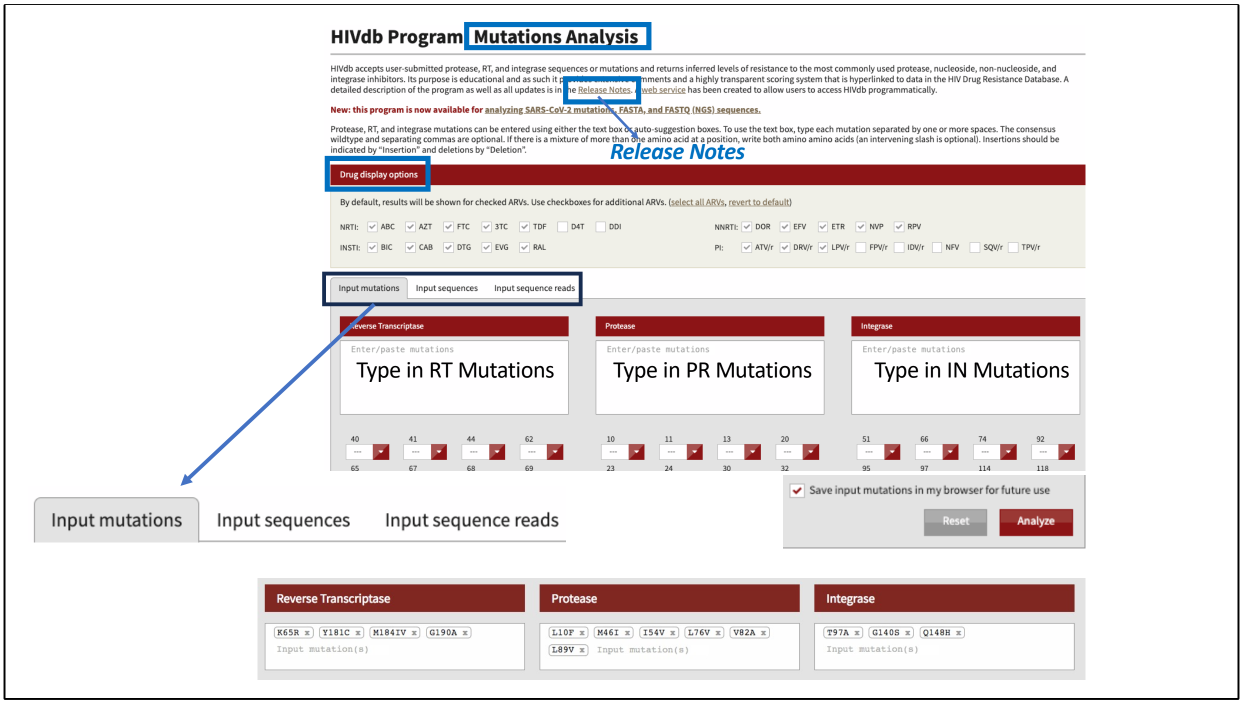
Genotype-treatment
ARV selection data comprising 217,525 protease, 227,074 RT, 35,861 integrase and 23,816 capsid HIV-1 virus sequences from 257,315 persons; 1,075 protease, 838 RT and 340 integrase HIV-2 virus sequences from 1,134 persons. In vitro selection data includes 1,111 HIV-1 in vitro selection data of PR, RT and IN.

Genotype-phenotype
Drug susceptibility data comprising 30,312 PI, 24,089 NRTI, 14,080 NNRTI and 5,370 INI susceptibility results from HIV-1 virus isolates

Genotype-clinical
Clinical outcome data comprising genotype, treatments, plasma HIV-1 RNA levels and CD4 counts from 15 clinical trials and >1500 Treatment-Change Episodes

References
2,105 references of genotype-treatment and/or genotype-phenotype data according to author-yr or according to year of addition to HIVDB.
3,763 Genbank submission sets according to author-yr and submission title, including 25 new submissions from Genbank release on 2023-12-15.
1,245 Genbank submission sets of capsid sequences according to author-yr and submission title from Genbank release on 2021-10-15. capsid mutation profile according to subtype..

This is the link to the HIVDB program from the HIVDB home page.



HIVdb Program Mutations Analysis

HIVdb accepts user-submitted protease, RT, and integrase sequences or mutations and returns inferred levels of resistance to the most commonly used protease, nucleoside, non-nucleoside, and integrase inhibitors. Its purpose is educational and as such it provides extensive comments and a highly transparent scoring system that is hyperlinked to data in the HIV Drug Resistance Database. A detailed description of the program as well as all updates is in the Release Notes. A web service has been created to allow users to access HIVdb programmatically.

New: this program is now available for analyzing SARS-CoV-2 mutations, FASTA, and FASTQ (NGS) sequences.

Protease, RT, and integrase mutations can be entered using either the text box or auto-suggestion boxes. To use the text box, type each mutation separated by one or more spaces. The consensus wildtype and separating commas are optional. If there is a mixture of more than one amino acid at a position, write both amino amino acids (an intervening slash is optional). Insertions should be indicated by “Insertion” and deletions by “Deletion”.

Drug display options

By default, results will be shown for checked ARVs. Use checkboxes for additional ARVs. (select all ARVs, revert to default)

NRTI: ☑ ABC ☑ AZT ☑ FTC ☑ 3TC ☑ TDF □ D4T □ DDI

NNRTI: ☑ DOR ☑ EFV ☑ ETR ☑ NVP ☑ RPV

INSTI: ☑ BIC ☑ CAB ☑ DTG ☑ EVG ☑ RAL

PI: ☑ ATV/r ☑ DRV/r ☑ LPV/r □ FPV/r □ IDV/r □ NFV □ SQV/r □ TPV/r

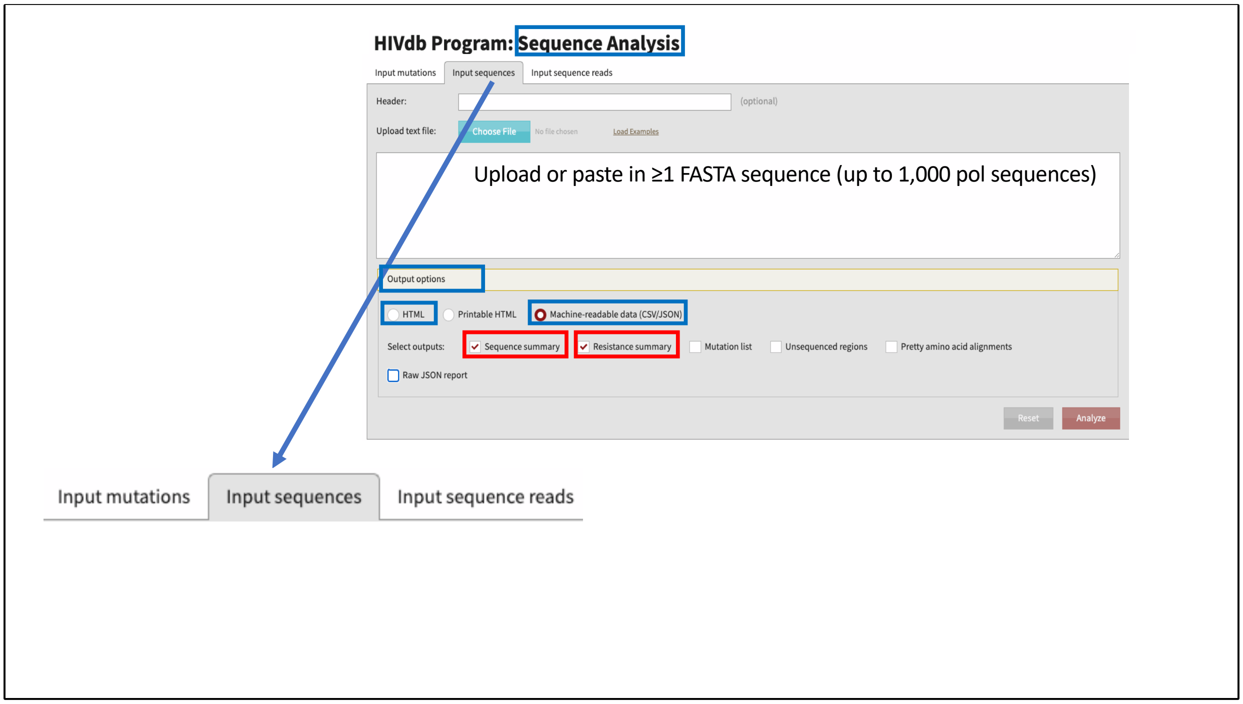
☑ Save input mutations in my browser for future use

Reverse Transcriptase: K65R Y181C M184IV G190A

Protease: L10F M46I I54V L76V V82A L89V

Integrase: T97A G140S Q148H

1. The sequence interpretation program accepts lists of mutations, one or more fasta sequences, and a fastq file.
2. Fasta sequences usually represent Sanger sequences of the consensus of the many reads from an NGS platform.
3. This slide shows how lists of mutation are analyzed.
4. I’ve highlighted several additional parts of the page: (i) The fact that the user selected the tab “Input mutations; (ii) a link to the Release Notes; and (iii) the Drug Display Options. By default, the program shows only those ARVs that are still being used.
5. Most often users type in RT, PR, and IN mutations in the three text boxes. However, it is also possible to use drop down menus to select each of the DRMs.
6. The text boxes allow the mutations to be entered in a variety of different formats.
7. By default the mutations entered are saved until the Reset button is clicked enabling the user to return to the page and modify the mutation list even after the interpretation is obtained.



HIVdb Program: Sequence Analysis

Input mutations | Input sequences | Input sequence reads

Header: (optional)

Upload text file: No file chosen

Upload or paste in ≥1 FASTA sequence (up to 1,000 pol sequences)

Output options

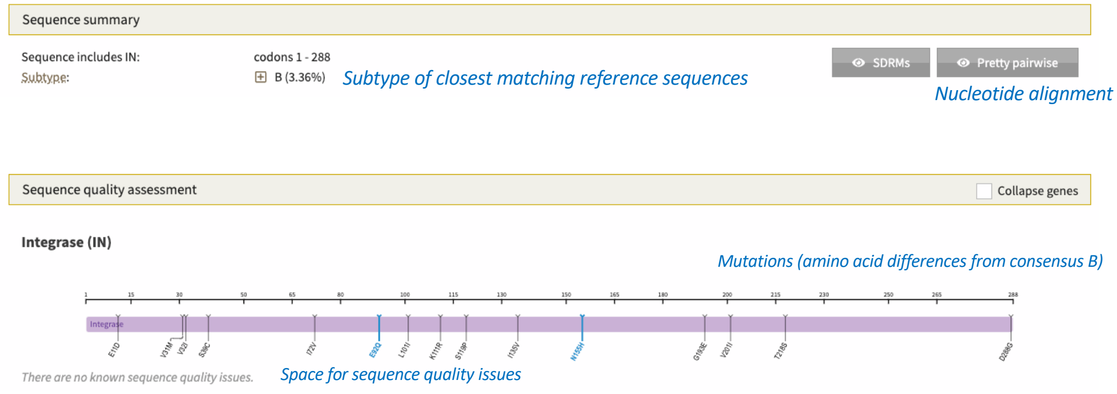
○ HTML ○ Printable HTML ● Machine-readable data (CSV/JSON)

Select outputs: ☑ Sequence summary ☑ Resistance summary □ Mutation list □ Unsequenced regions □ Pretty amino acid alignments

□ Raw JSON report

1. This slide shows how to either paste in or upload a list of fasta sequences. At least 1000 FASTA complete pol sequences can be pasted in.
2. The slide also shows that the two main output options are HTML or a machine-readable file – usually csv file.
3. Finally among the machine readable formats, the most useful are the Sequence summary and the Resistance Summary.
4. The HTML output is fine if a small number of sequences is entered but machine readable csv files are much more useful if many sequences are entered.

HTML Output: Sequence Summary, Subtype, Quality Control


Sequence summary

Sequence includes IN: codons 1 - 288

Subtype: ⊞ B (3.36%)

Sequence quality assessment

□ Collapse genes

Integrase (IN)

There are no known sequence quality issues.

Subtype of closest matching reference sequences

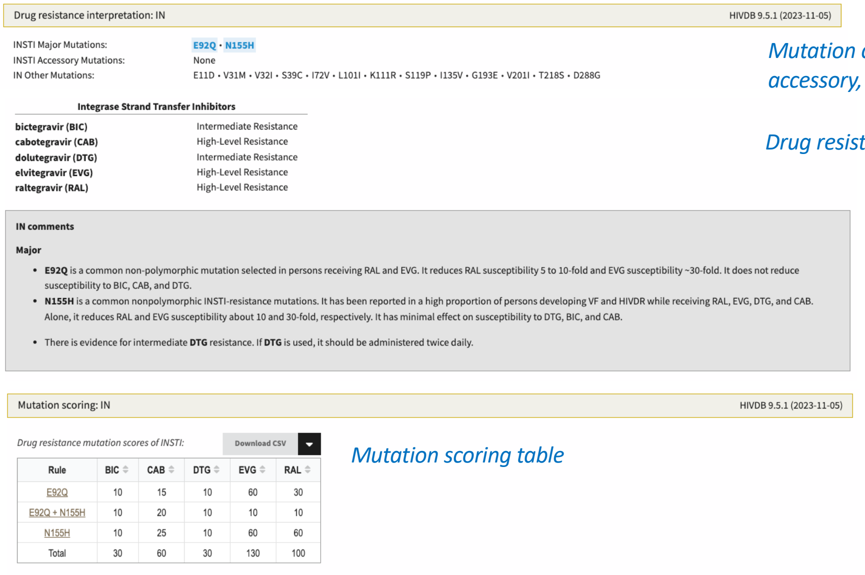
Nucleotide alignment

Mutations (amino acid differences from consensus B)

Space for sequence quality issues

1. This slide and the next one will review the HTML output of the interpretation program.
2. This is the only way to review the results when a mutation list is entered. It is also a common way to the review the results when a small number of sequences are entered.
3. The very top part of the result shows the region that was sequenced, the closest matching reference sequence and its genetic distance from the submitted sequence, and a link to the nucleotide alignment.
4. This example shows that the sequence contained the complete integrase gene.
5. By clicking on the plus sign, the top 10 closest matching reference sequences can be viewed
6. The 2nd part of the output contains a figure showing the mutations defined as amino acid differences from consensus B. Drug resistance mutations and mutations that likely reflect sequence quality control issues are indicated by different colors.
7. Beneath the figure there is a textual summary of sequence quality control issues when present. These include stop codons, frame shifts, mutations suggesting that APOBEC-mediated G-to-A hypermutation is present, and highly unusual mutations.

HTML Drug Resistance Interpretation



Drug resistance interpretation: IN | HIVDB 9.5.1 (2023-11-05)

INSTI Major Mutations: E92Q • N155H
INSTI Accessory Mutations: None
IN Other Mutations: E11D • V31M • V32I • S39C • I72V • L101I • K111R • S119P • I135V • G193E • V201I • T218S • D288G

Integrase Strand Transfer Inhibitors	
bictegravir (BIC)	Intermediate Resistance
cabotegravir (CAB)	High-Level Resistance
dolutegravir (DTG)	Intermediate Resistance
elvitegravir (EVG)	High-Level Resistance
raltegravir (RAL)	High-Level Resistance

IN comments

Major

- **E92Q** is a common non-polymorphic mutation selected in persons receiving RAL and EVG. It reduces RAL susceptibility 5 to 10-fold and EVG susceptibility ~30-fold. It does not reduce susceptibility to BIC, CAB, and DTG.
- **N155H** is a common nonpolymorphic INSTI-resistance mutations. It has been reported in a high proportion of persons developing VF and HIVDR while receiving RAL, EVG, DTG, and CAB. Alone, it reduces RAL and EVG susceptibility about 10 and 30-fold, respectively. It has minimal effect on susceptibility to DTG, BIC, and CAB.

- There is evidence for intermediate **DTG** resistance. If **DTG** is used, it should be administered twice daily.

Mutation scoring: IN | HIVDB 9.5.1 (2023-11-05)

Drug resistance mutation scores of INSTI:

Rule	BIC	CAB	DTG	EVG	RAL
E92Q	10	15	10	60	30
E92Q + N155H	10	20	10	10	10
N155H	10	25	10	60	60
Total	30	60	30	130	100

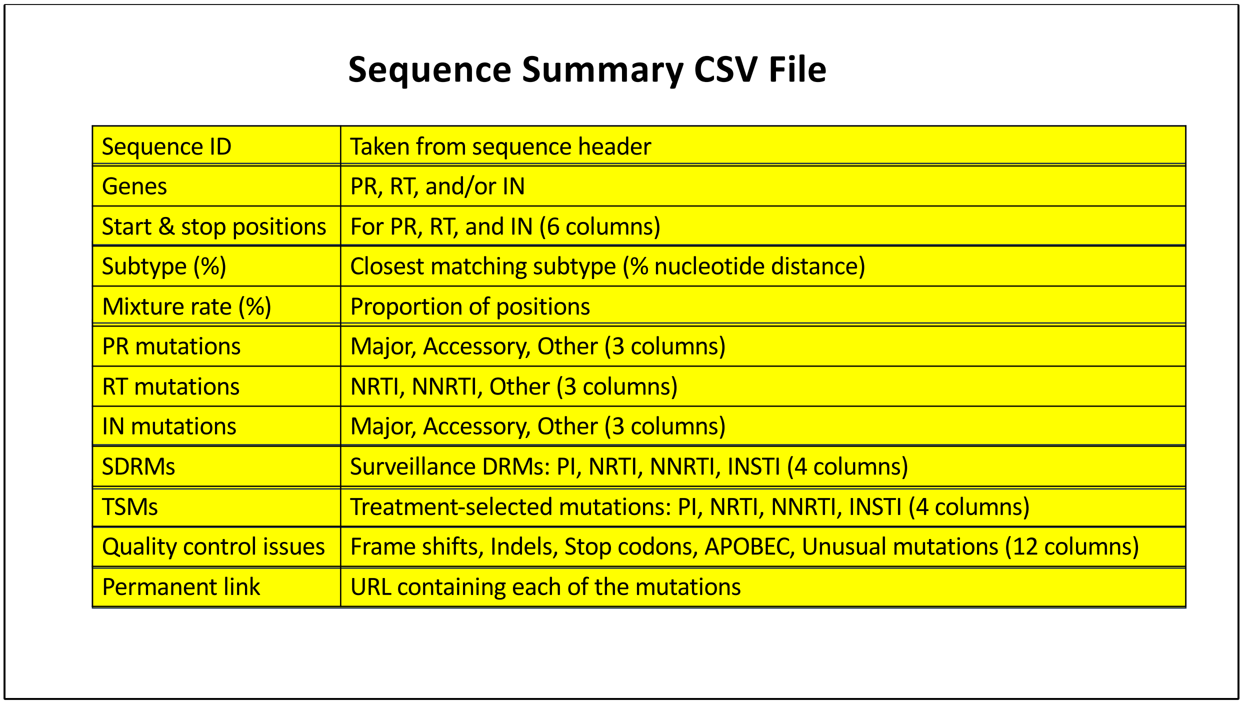
Mutation classification: major, accessory, other

Drug resistance levels

Mutation comments

Mutation scoring table

1. The remainder of the output includes the drug resistance interpretation for each gene that was sequenced.
2. For integrase, protease, and capsid, the mutations are divided into major, accessory, and other. Other includes mutations that do not receive penalty scores.
3. For RT, the mutations are divided into NRTI, NNRTI, and other.
4. The next part of the output contains the drug-resistance interpretation based on the five levels shown previously and a list of mutation comments.
5. The final part of the output for each gene contains the mutation scoring table, which shows how the individual and combination DRM penalties are added up to yield the total score that is used to assign one of the 5 drug levels.



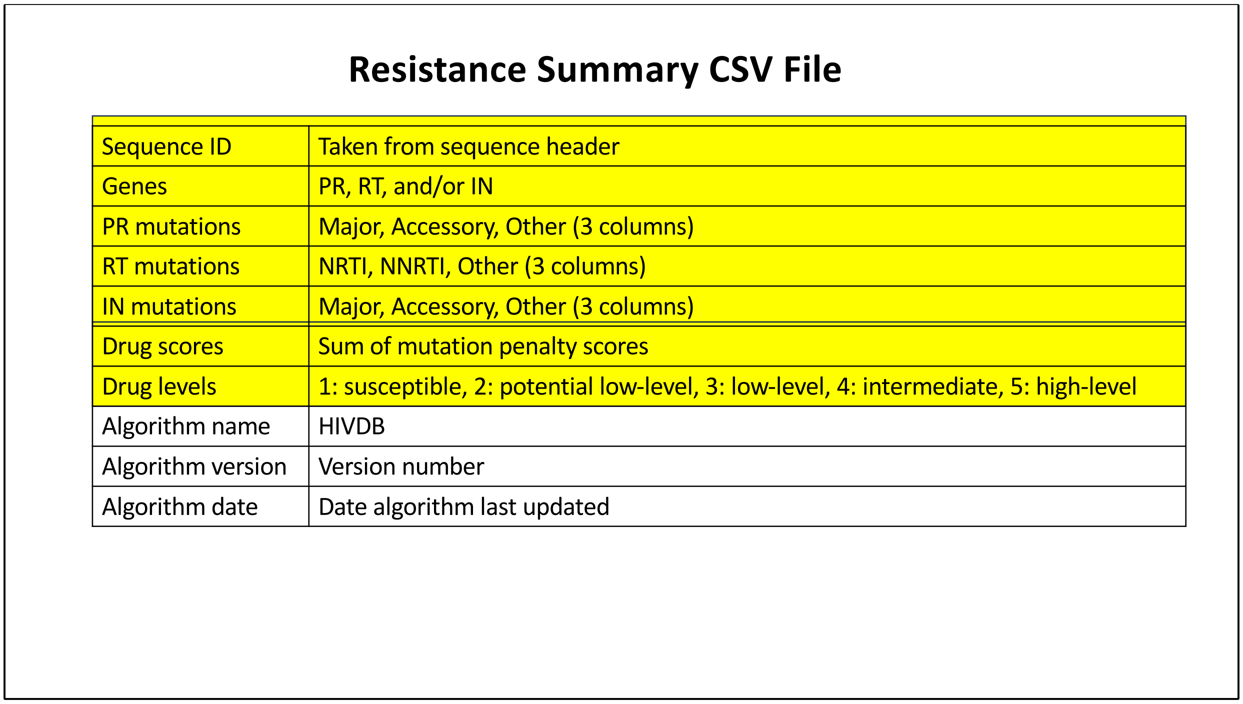
Sequence Summary CSV File

Sequence ID	Taken from sequence header
Genes	PR, RT, and/or IN
Start & stop positions	For PR, RT, and IN (6 columns)
Subtype (%)	Closest matching subtype (% nucleotide distance)
Mixture rate (%)	Proportion of positions
PR mutations	Major, Accessory, Other (3 columns)
RT mutations	NRTI, NNRTI, Other (3 columns)
IN mutations	Major, Accessory, Other (3 columns)
SDRMs	Surveillance DRMs: PI, NRTI, NNRTI, INSTI (4 columns)
TSMs	Treatment-selected mutations: PI, NRTI, NNRTI, INSTI (4 columns)
Quality control issues	Frame shifts, Indels, Stop codons, APOBEC, Unusual mutations (12 columns)
Permanent link	URL containing each of the mutations

1. If a large number of sequences are entered, then it makes more sense to select one of the spreadsheet output options of which the two most common are the ”Sequence Summary” and “Resistance Summary” file.
2. This figure describes the output of the Sequence Summary file.
3. Overall, the file contains one row for each sequence and 40 columns.
4. The first column contains the unique Sequence ID and the next 7 indicate which genes were sequenced and the range of positions that were sequenced.
5. There is a column with the genetic distance to the closest matching subtype and a column indicating the proportion of positions containing a mixture of more more than one nucleotide.
6. In newly diagnosed untreated persons, the proportion of positions with mixtures has been shown to correlate with the duration of infection.
7. The next 9 columns list the 3 categories of mutations for PR, RT, and IN.
8. The next 4 columns list the DRMs which have been used for surveillance of transmitted drug resistance.
9. The surveillance DRMs or SDRMs overlap a lot with the DRMs shown in the preceding columns. However, they differ in that none are polymorphic and that none are extremely rare.
10. In contrast the treatment-selected mutations or TSMs are nonpolymorphic

mutations that are mutations without DRM penalty scores that are significantly more common in treated persons than untreated persons.

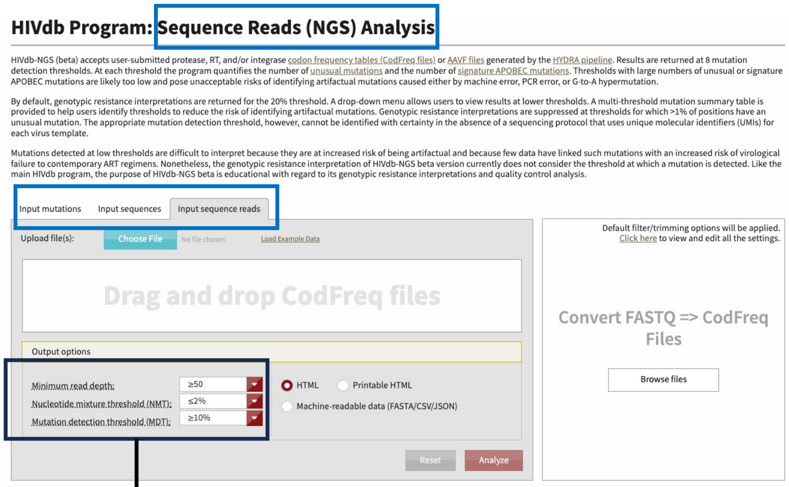
11. The TSMs are generally extremely rare. They generally only occur in combination with other well-established DRMs. As a result, they have not been well studied and therefore have not been assigned penalty scores.
12. There are 12 columns highlighting QC issues.
13. The last column is a URL containing each of the mutations which allows the HTML output to be viewed.



Resistance Summary CSV File

Sequence ID	Taken from sequence header
Genes	PR, RT, and/or IN
PR mutations	Major, Accessory, Other (3 columns)
RT mutations	NRTI, NNRTI, Other (3 columns)
IN mutations	Major, Accessory, Other (3 columns)
Drug scores	Sum of mutation penalty scores
Drug levels	1: susceptible, 2: potential low-level, 3: low-level, 4: intermediate, 5: high-level
Algorithm name	HIVDB
Algorithm version	Version number
Algorithm date	Date algorithm last updated

1. This slide summarizes the columns shown in the Resistance Summary spreadsheet.
2. In addition to having the 9 columns showing each of the PR, RT, and IN mutations, it also has columns showing total score and resistance levels for the drugs belonging to each drug class.



HIVdb Program: Sequence Reads (NGS) Analysis

HIVdb-NGS (beta) accepts user-submitted protease, RT, and/or integrase codon frequency tables (CodFreq files) or AAVF files generated by the HYDRA pipeline. Results are returned at 8 mutation detection thresholds. At each threshold the program quantifies the number of unusual mutations and the number of signature APOBEC mutations. Thresholds with large numbers of unusual or signature APOBEC mutations are likely too low and pose unacceptable risks of identifying artifactual mutations caused either by machine error, PCR error, or G-to-A hypermutation.

By default, genotypic resistance interpretations are returned for the 20% threshold. A drop-down menu allows users to view results at lower thresholds. A multi-threshold mutation summary table is provided to help users identify thresholds to reduce the risk of identifying artifactual mutations. Genotypic resistance interpretations are suppressed at thresholds for which >1% of positions have an unusual mutation. The appropriate mutation detection threshold, however, cannot be identified with certainty in the absence of a sequencing protocol that uses unique molecular identifiers (UMIs) for each virus template.

Mutations detected at low thresholds are difficult to interpret because they are at increased risk of being artifactual and because few data have linked such mutations with an increased risk of virological failure to contemporary ART regimens. Nonetheless, the genotypic resistance interpretation of HIVdb-NGS beta version currently does not consider the threshold at which a mutation is detected. Like the main HIVdb program, the purpose of HIVdb-NGS beta is educational with regard to its genotypic resistance interpretations and quality control analysis.

Upload file(s): No file chosen

Default filter/trimming options will be applied.
Click here to view and edit all the settings.

Output options

Minimum read depth: ≥50
Nucleotide mixture threshold (NMT): ≤2%
Mutation detection threshold (MDT): ≥10%

● HTML ○ Printable HTML
○ Machine-readable data (FASTA/CSV/JSON)

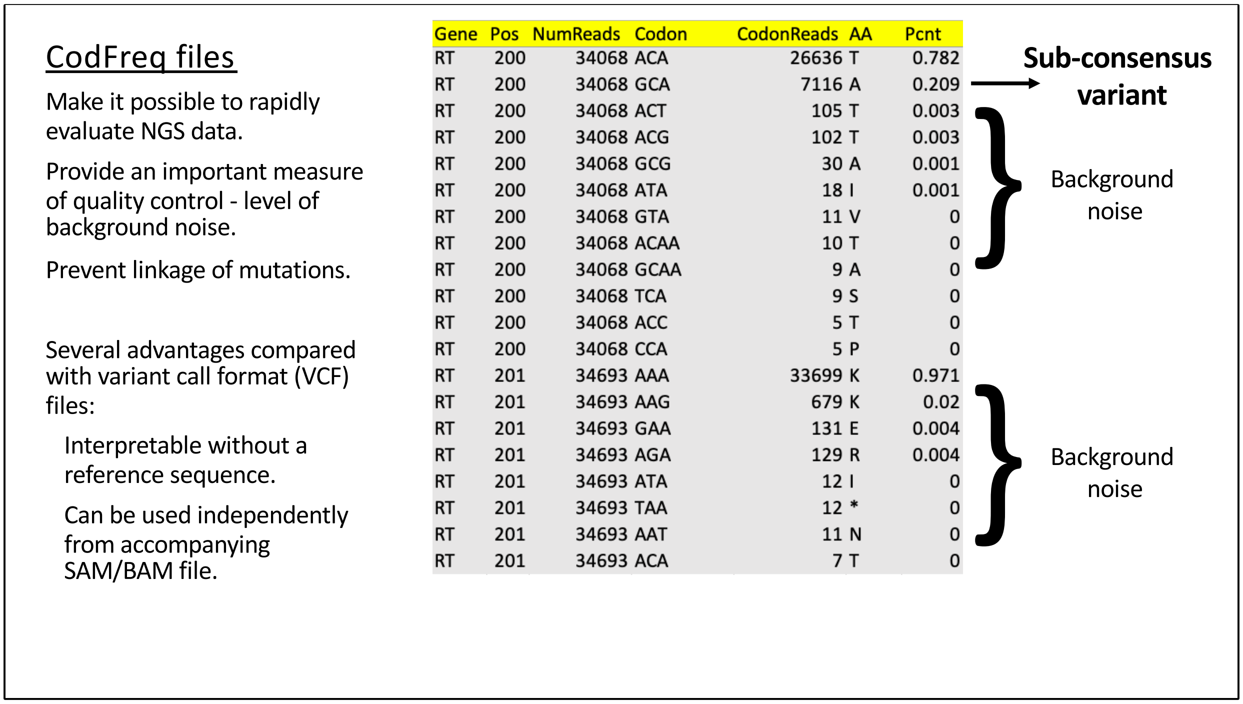
Interpretation of CodFreq files takes 2-3 seconds.

Conversion of FASTQ files to CodFreq files takes about 2-5 minutes. It can take place on our website or the user's computer.

Select minimum read depth and threshold for detecting sub-consensus mutations

1. By clicking on the ”Input sequence reads tab” it is possible to analyze NGS data.
2. This requires two steps. The first requires converting one or more FASTQ files to a format we developed called a codon frequency or CodFreq file.
3. The following slide will describe CodFreq files in detail.
4. This conversion can take place by dragging the FASTQ files to the box at the right or by uploading one or more FASTQ files.
5. If two files have the same prefix, they are considered to be paired files.
6. The conversion of a FASTQ file to a CodFreq file takes 2-5 minutes depending on the size of the file. It can be done on our website or on the user's computer.
7. The interpretation of a CodFreq file takes 2-3 seconds.
8. The user can also select the minimum reads required for a region to be considered sequenced, the mutation detection threshold which is the minimum proportion of reads containing the mutation for it to be considered present, and the maximum allowable proportion of mixed nucleotides.
9. By default the minimum number of reads is set at 50 and the mutation detection threshold is set at 10%.

OXFORD NANOPORE HAS one file
Ion torrent file
Illumina 1 or 2. If multiple are added the program will combine the paired reads.



CodFreq files

Make it possible to rapidly evaluate NGS data.

Provide an important measure of quality control - level of background noise.

Prevent linkage of mutations.

Several advantages compared with variant call format (VCF) files:

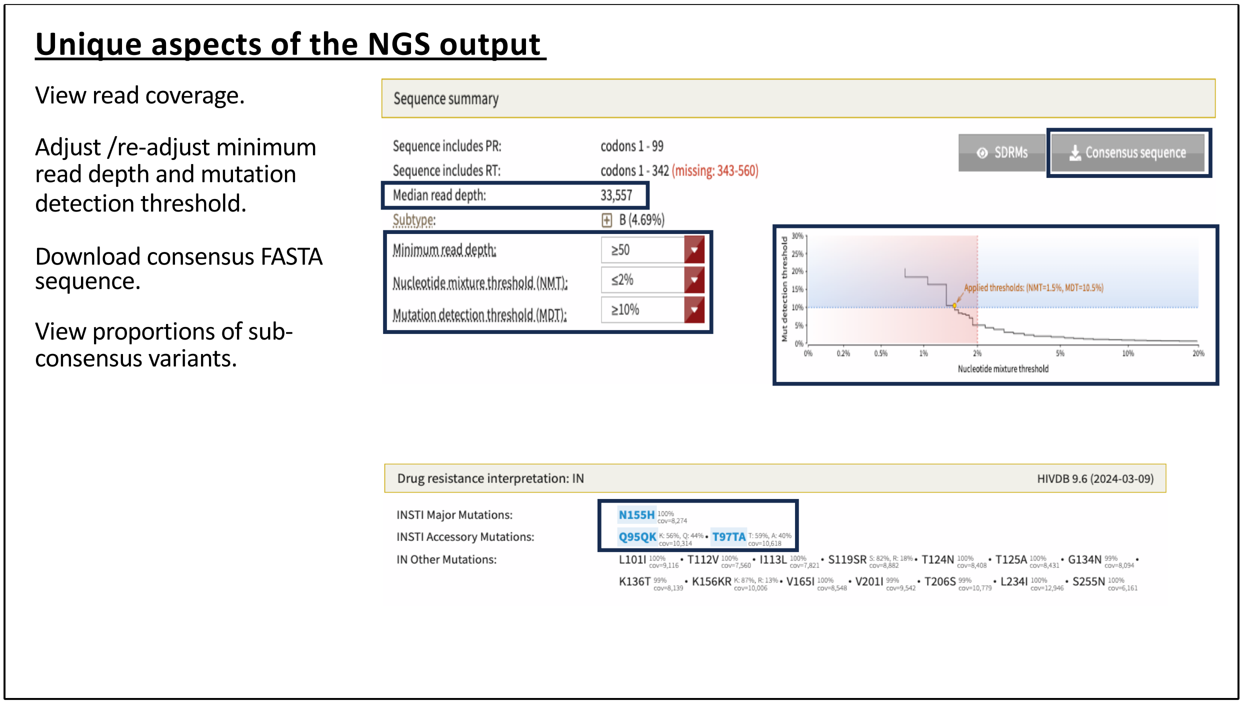
- Interpretable without a reference sequence.
- Can be used independently from accompanying SAM/BAM file.

Gene	Pos	NumReads	Codon	CodonReads	AA	Pcnt
RT	200	34068	ACA	26636	T	0.782
RT	200	34068	GCA	7116	A	0.209
RT	200	34068	ACT	105	T	0.003
RT	200	34068	ACG	102	T	0.003
RT	200	34068	GCG	30	A	0.001
RT	200	34068	ATA	18	I	0.001
RT	200	34068	GTA	11	V	0
RT	200	34068	ACAA	10	T	0
RT	200	34068	GCAA	9	A	0
RT	200	34068	TCA	9	S	0
RT	200	34068	ACC	5	T	0
RT	200	34068	CCA	5	P	0
RT	201	34693	AAA	33699	K	0.971
RT	201	34693	AAG	679	K	0.02
RT	201	34693	GAA	131	E	0.004
RT	201	34693	AGA	129	R	0.004
RT	201	34693	ATA	12	I	0
RT	201	34693	TAA	12	*	0
RT	201	34693	AAT	11	N	0
RT	201	34693	ACA	7	T	0

1. CodFreq files contain tables with 7 columns in which the frequency of each codon at each position is indicated.
2. In this example the different codons present at positions 200 and 201 are shown.
3. Overall, there are approximately 34,000 reads encompassing these two positions.
4. Most of the codons are present at very low levels and likely represent PCR or machine errors.
5. The mutation T200A is present in about 21% of reads. Nearly all of the remaining mutations – with the possible silent mutation present in 2% of reads at position 201 – likely represent background noise.
6. There are three main advantages of working with codon frequency files. First, they are much smaller than FASTQ files. Second, drug resistance interpretations can be generated very rapidly.
7. Third, the distribution in the frequency of non-consensus codons provides an important measure of quality control. For example, if the level of background noise is high, it is important to be very conservative when setting the mutation detection threshold. Otherwise many of the variants will be sequence artifacts.
8. The one disadvantage of codon frequency files is that it is not possible to know whether sub-consensus mutations at different positions occur on the same reads. This is a minor disadvantage because it is never possible to be sure of linkage

unless single genome amplification was performed prior to sequencing and because linkage of mutations is currently not considered during genotypic resistance interpretation.

9. CodFreq files have two advantages over the established variant call format or VSF files. They are interpretable without a reference sequence and they can be used independently from an accompanying SAM or BAM file.



Unique aspects of the NGS output

View read coverage.

Adjust /re-adjust minimum read depth and mutation detection threshold.

Download consensus FASTA sequence.

View proportions of sub-consensus variants.

Sequence summary

Sequence includes PR: codons 1 - 99

Sequence includes RT: codons 1 - 342 (missing: 343-560)

Median read depth: 33,557

Subtype: ⊞ B (4.69%)

Minimum read depth: ≥50

Nucleotide mixture threshold (NMT): ≤2%

Mutation detection threshold (MDT): ≥10%

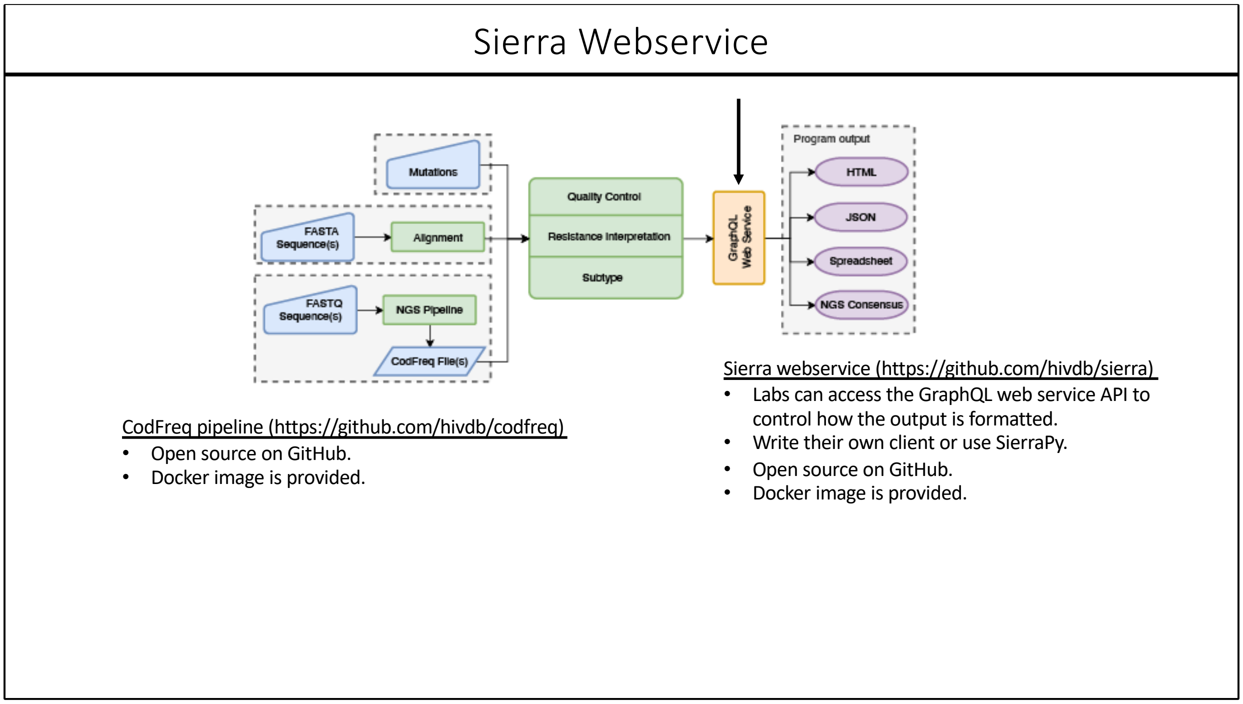
Drug resistance interpretation: IN HIVDB 9.6 (2024-03-09)

INSTI Major Mutations: N155H 100% cov=8,274

INSTI Accessory Mutations: Q95QK K: 56%, Q: 44% cov=10,314 • T97TA T: 59%, A: 40% cov=10,618

IN Other Mutations: L101I 100% cov=9,116 • T112V 100% cov=7,560 • I113L 100% cov=7,821 • S119SR S: 82%, R: 18% cov=8,882 • T124N 100% cov=8,408 • T125A 100% cov=8,431 • G134N 99% cov=8,094 • K136T 99% cov=8,139 • K156KR K: 87%, R: 13% cov=10,006 • V165I 100% cov=8,548 • V201I 99% cov=9,542 • T206S 99% cov=10,779 • L234I 100% cov=12,946 • S255N 100% cov=6,161

1. This slide shows those aspects of the output that are unique to NGS sequences.
2. The median read depth is shown and those parts of the sequence that have fewer reads than the minimum read depth are also indicated (although I am not showing those “warnings” on this slide).
3. After the interpretation is obtained, it is possible to reset the minimum read depth and two adjust two different thresholds: the proportion of positions allowed to contain a mixture of nucleotides and the mutation detection threshold above which a mutation has to be present in order for it to be considered to be present.
4. Changes to these three parameters will be immediately reflected in the consensus sequence in that a lower threshold will result in more nucleotide ambiguities and the detection of more low-abundance mutations.
5. The list of reported mutations may change and if they include DRMs the overall report will change.
6. The output also indicates the read coverage at each position and the proportion of reads for each mutation.
7. Finally, this figure can help select a reasonable mutation detection threshold.



Sierra Webservice

CodFreq pipeline (https://github.com/hivdb/codfreq)

- Open source on GitHub.
- Docker image is provided.

Sierra webservice (https://github.com/hivdb/sierra)

- Labs can access the GraphQL web service API to control how the output is formatted.
- Write their own client or use SierraPy.
- Open source on GitHub.
- Docker image is provided.

•1. This slide shows an overall schematic of the drug resistance interpretation programs.

•2. Users can submit a list of mutations, one or more FASTA sequences, or one or more CodFreq files.

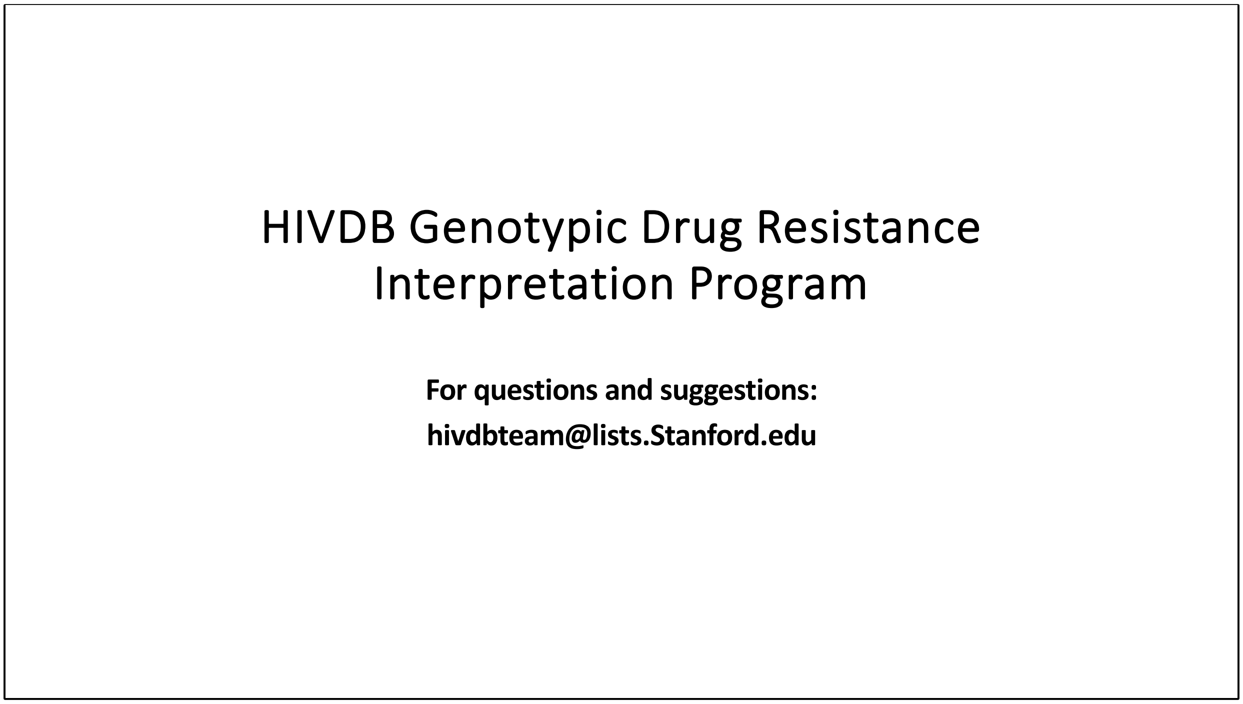
•3. CodFreq files can be generated from FASTQ files on our web site. The process can also be automated using opensource code. The code is available as a Docker image so that it can be run locally.

•4. All of the output is delivered to the user using a GraphQL webservice which we use to deliver the output on our website in a variety of different formats.

•5. However, if users want to control the process of interacting with the drug resistance interpretation program and of formatting the output, they can access the GraphQL web service through its API.

•6. They can write their own client or use a client we developed called SierraPy.

•7. They can also run the entire program locally using the Sierra Docker image.



HIVDB Genotypic Drug Resistance Interpretation Program

For questions and suggestions:

hivdbteam@lists.Stanford.edu