COLLEGE of AMERICAN PATHOLOGISTS

Preventing Common Next-Generation Sequencing Testing Errors

Common Laboratory Errors in Next-Generation Sequencing Assays and How to Mitigate Them

Valentina Nardi, MD, FCAP
Assistant Pathologist
Massachusetts General Hospital
Boston, MA, US

December 6, 2022

Presenter: Valentina Nardi, MD, FCAP

- Associate Pathologist at Massachusetts General Hospital
  - Molecular genetic pathology, Hematopathology and Gastrointestinal pathology
- Associate Professor at Harvard Medical School
- American Board Certifications
  - Anatomic Pathology, American Board of Pathology
  - Hematology, American Board of Pathology
  - Molecular Genetic Pathology, American Board of Medical Genetics
- Involvement with the College of American Pathologists (CAP)
  - Member of Molecular Oncology Committee
  - Fellow

Disclosures

- Loxo Oncology, SAB participant
- Novartis: speaker honorarium

Objectives

- Understand the recurrent weaknesses of laboratories performing Next-Generation Sequencing (NGS) assays using NGS Proficiency Testing/External Quality Assessment (PT/EQA) data.
- Offer suggestions on how to mitigate the most common causes of unacceptable results on PT/EQA and improve sensitivity and specificity of NGS assays.

Scenario One

60-year-old male with acute leukemia (>95% tumor).

NGS results off bone marrow are consistent with a female patient (no Y chromosome SNPs detected).

How do you explain it?

- ✓ A. Sample swap
- ✓ B. Poor sequencing quality
- ✓ C. Transgender patient
- ✓ D. Bone marrow recipient with female donor
- ✓ E. Loss of Y chromosome in the leukemia

Scenario Two

50-year-old, Asian woman who never smoked with NSCLC.

Targeted DNA and RNA sequencing detect no variants.

How do you explain it?

A. Driver not covered by the assays

B. Sample swap

C. Variant filtered out by the pipeline

D. Insufficient tumor cellularity

✓ E. All of the above

Scenario Three

An oncologist comes to you as an outside laboratory detected a KRAS G12C mutation in his patient's NSCLC, but your laboratory does not.

How do you reconcile the results?

Which one is the Correct Annotation?

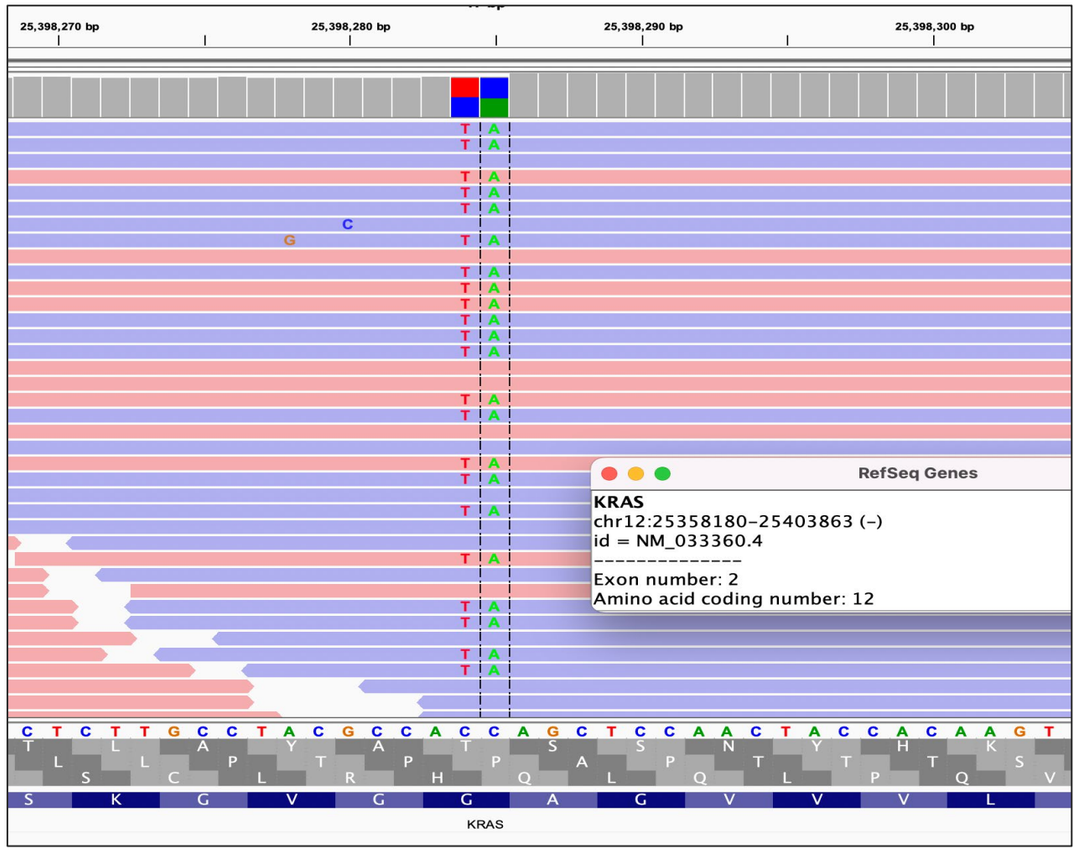
A. KRAS c.34C>A p.Gly12Cys &
   KRAS c.35C>T, p.Gly12Asp

B. KRAS c.34G>T p.Gly12Cys &
   KRAS c.35G>A, p.Gly12Asp

C. KRAS c.34_35GG>TA, p. Gly12Tyr

D. KRAS c.34_35delGGinsTA, p. Gly12Tyr

E. Unsure



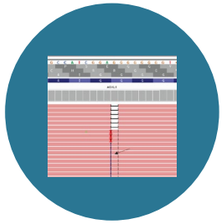
How Common are Errors in NGS PT Surveys? What is the Size of the Problem?

- Really small!
  - Somatic PT surveys: 95.9% accuracy (2016–2019)
  - Germline PT surveys: 97.8% accuracy (2015–2019)

Three type of errors

1. Pre-analytical

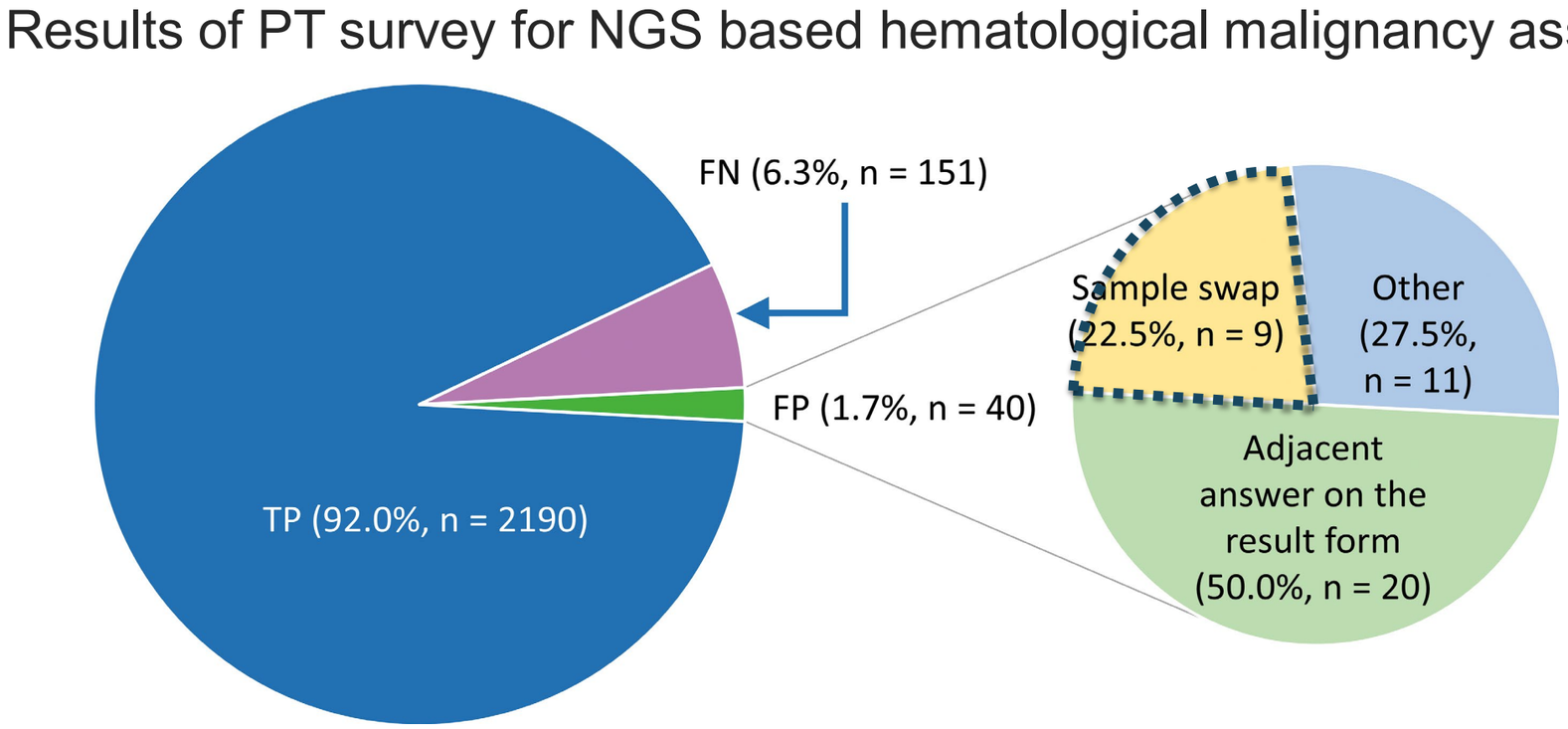
2. Analytical
   - False positive
   - False negative
   - Errors in annotation



3. Post-analytical

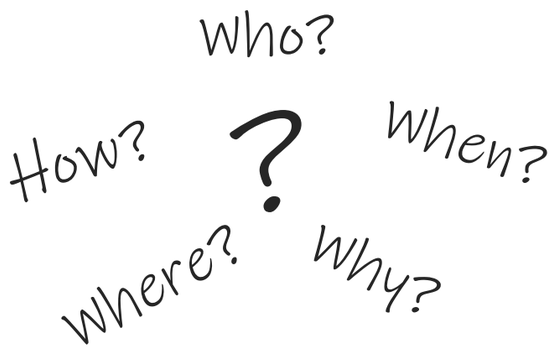
1. Pre-analytical Errors:

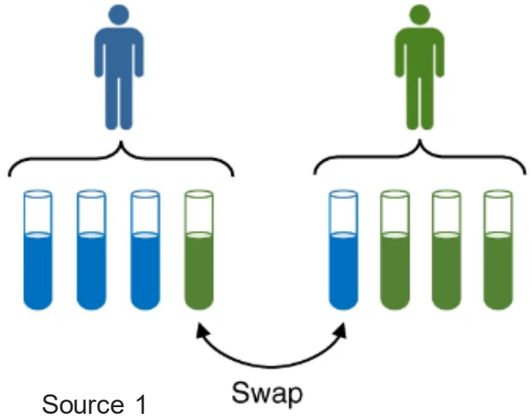
A. Sample mix up, leading to False Positive (FP) and False Negative (FN) results



Reference: Keegan A, et al. *Arch Pathol Lab Med*. (2020)

Sample Swaps





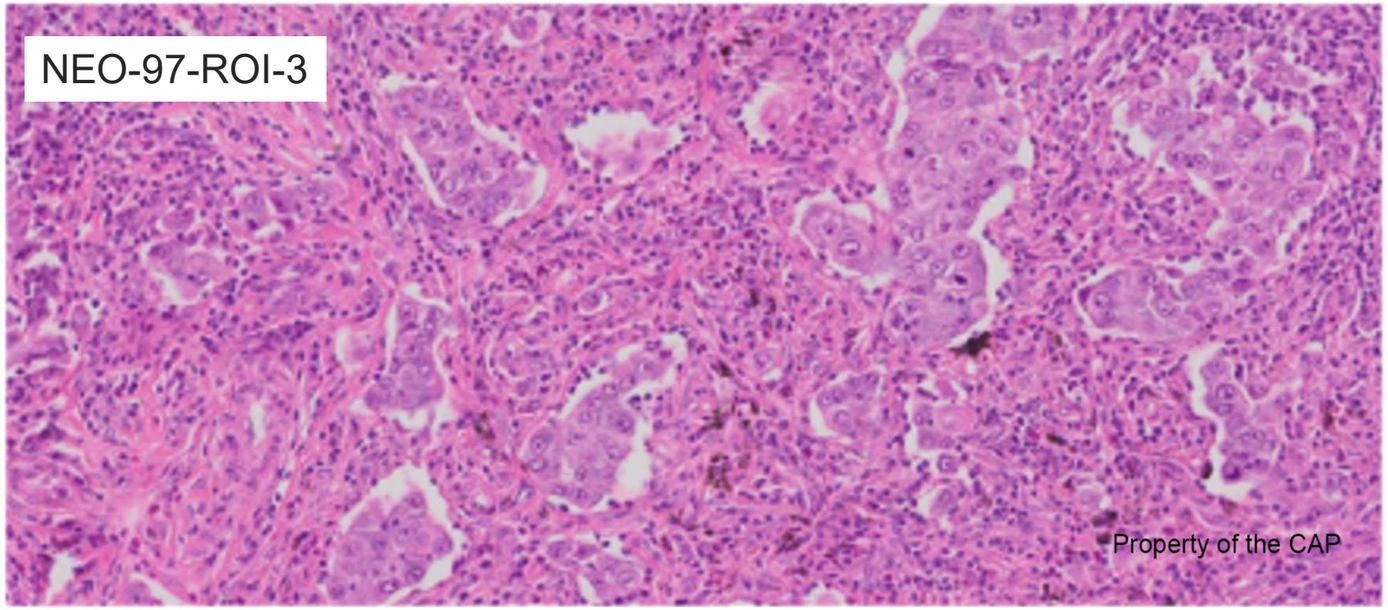
- Mislabeling of samples
- Mislabeling of samples sheets
- Rotated sample plate
- Informatic error

Sample Swaps' "Remedies"

- ✓ Avoid multiple patient specimens in the active work area at the same time
- ✓ Label 1 specimen at a time before proceeding to the next one
- ✓ Have a second person checking the labeling of tubes
- ✓ Verify clinical samples according to sex by checking non-autosomal markers
- ✓ Adding unique DNA control sequences to samples (Moore R., et al, *J Mol Diagn.* 2020)

1. Pre-analytical Errors:

B. Errors in assessing neoplastic cellularity/specimen adequacy



Lung ACA, neoplastic cellularity estimate:

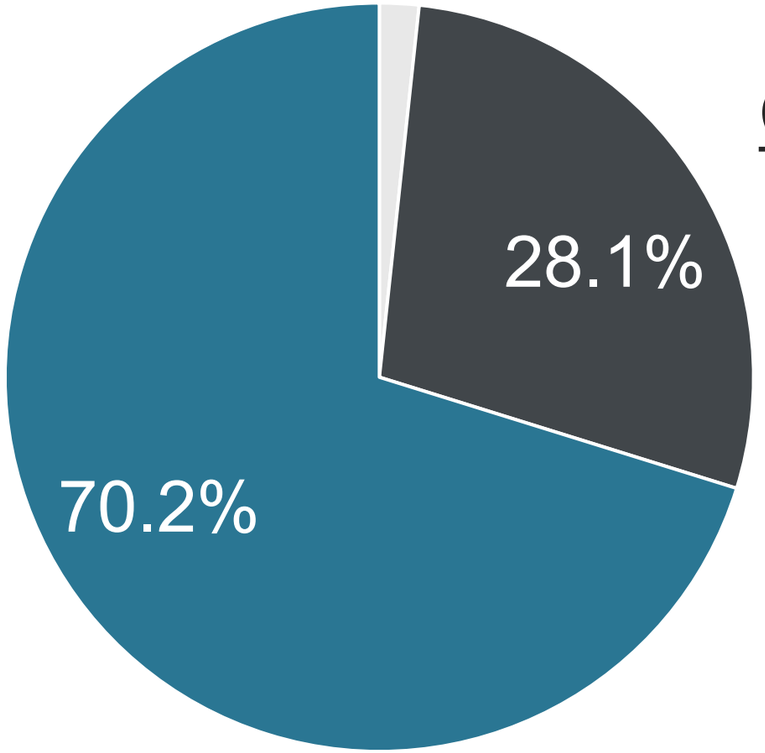
1) 10%
2) 20%
3) 30%
4) 40%
5) 50%

Table 7. Summary of Performance Challenge Images and Participant Performance

Image Information				Participant Performance			Adequacy Assessment		
				Neoplastic Cellularity, n = 54 for All Cases, %					
Case	Adenocarcinoma	Unique Histologic Feature	Criterion Standard, %	Mean (SD), CV %	Median	Minimum–Maximum	Assay Threshold, %	Adequacy Response	No. (%)
NEO-97 ROI-3	Lung	Abundant lymphocytes	9.3	22.3 (10.3), 46.2	20.0	5.0–65.0	30	Yes	11 (20.4)
								No	43 (79.6)

Reference: Devereaux KA. et al, *Arch Pathol Lab Med.* (2022)

Neoplastic Cellularity Assessment Method



Cell Area Method

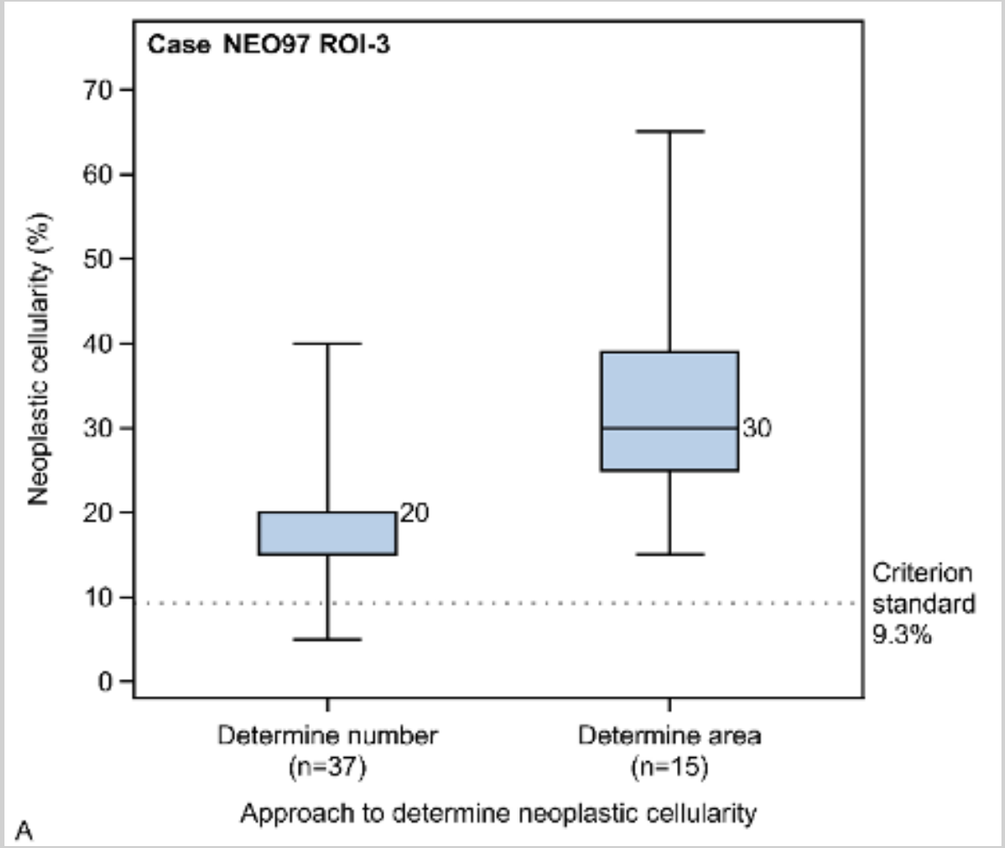
Area of neoplastic cells compared to overall area of tissue

Cell Number Method

Number of neoplastic cells compared to overall number of cells

Reference: Devereaux KA. et al, *Arch Pathol Lab Med*. (2022)
Slide Used with Permission from Joel T. Moncur, MS, MD, PhD, FCAP

$$\text{Neoplastic cellularity} = \frac{\text{number of neoplastic cells}}{\text{overall number of cells}}$$



- The number of cells or nuclei in a sample directly correlates with DNA content.
- Laboratories should define neoplastic cellularity as the number of neoplastic cells or nuclei compared to the overall number of cells.

Reference: Devereaux KA. et al, *Arch Pathol Lab Med. (*2022)

Neoplastic Cellularity Assessment Practices Do's and Don'ts

Table 2. Neoplastic Cellularity Assessment Practices

Practice	No. (%)
Certification of pathologist who routinely assesses the neoplastic cellularity	57
AP or AP/CP board-certified (may include fellows) (CAP accreditation requirement!!)	48 (84.2)
CP-only board-certified (may include fellows)	6 (10.5)
Non–board-certified anatomic pathologist only or AP/CP resident and/or fellow	3 (5.3)
Type of slide and/or image used to assess neoplastic cellularity (multiple responses allowed)	57
Glass slide	52 (91.2)
Digital slide	9 (15.8)
Static image of a slide	1 (1.8)
Approach to determining neoplastic cellularity	57
Determine the number of neoplastic cells or nuclei and compare to the overall number of cells	40 (70.2)
Determine the area of neoplastic cells and compare to the overall area of tissue	16 (28.1)
Determine by both number and area approaches	1 (1.8)
Method used to routinely assess neoplastic cellularity	57
Cellularity is estimated without counting cells	47 (82.5)
Cells are manually counted	10 (17.5)
Average time to determine the percentage of neoplastic cells for a single case using laboratory's routine method	57
<30 s	4 (7.0)
30 s to 3 min	42 (73.7)
>3 to 10 min	7 (12.3)
>10 min to 1 h	3 (5.3)
>1 h	1 (1.8)

Abbreviations: AP, anatomic pathology; CP, clinical pathology.

Reference: Devereaux KA. et al, *Arch Pathol Lab Med. (*2022)

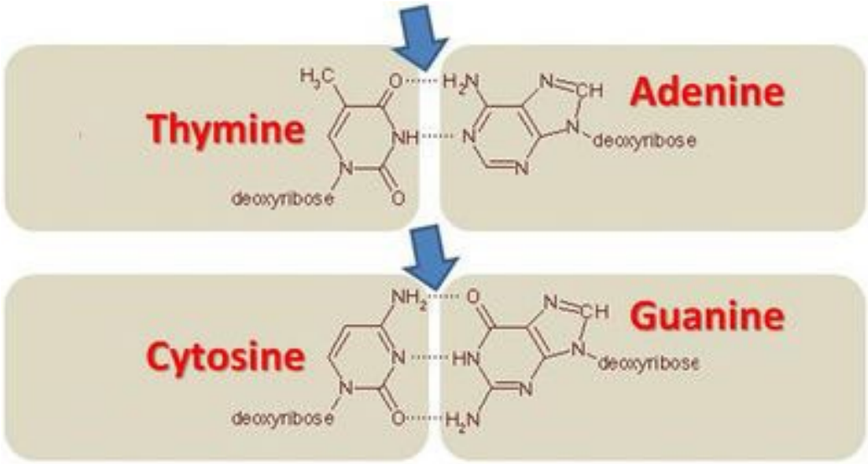
2. Analytical Errors:

A. False Negative (FN) errors – Variants in difficult-to-sequence regions with high GC content

"GC rich" = ~60% of the bases are either cytosine (C) or guanine (G)

Problems:

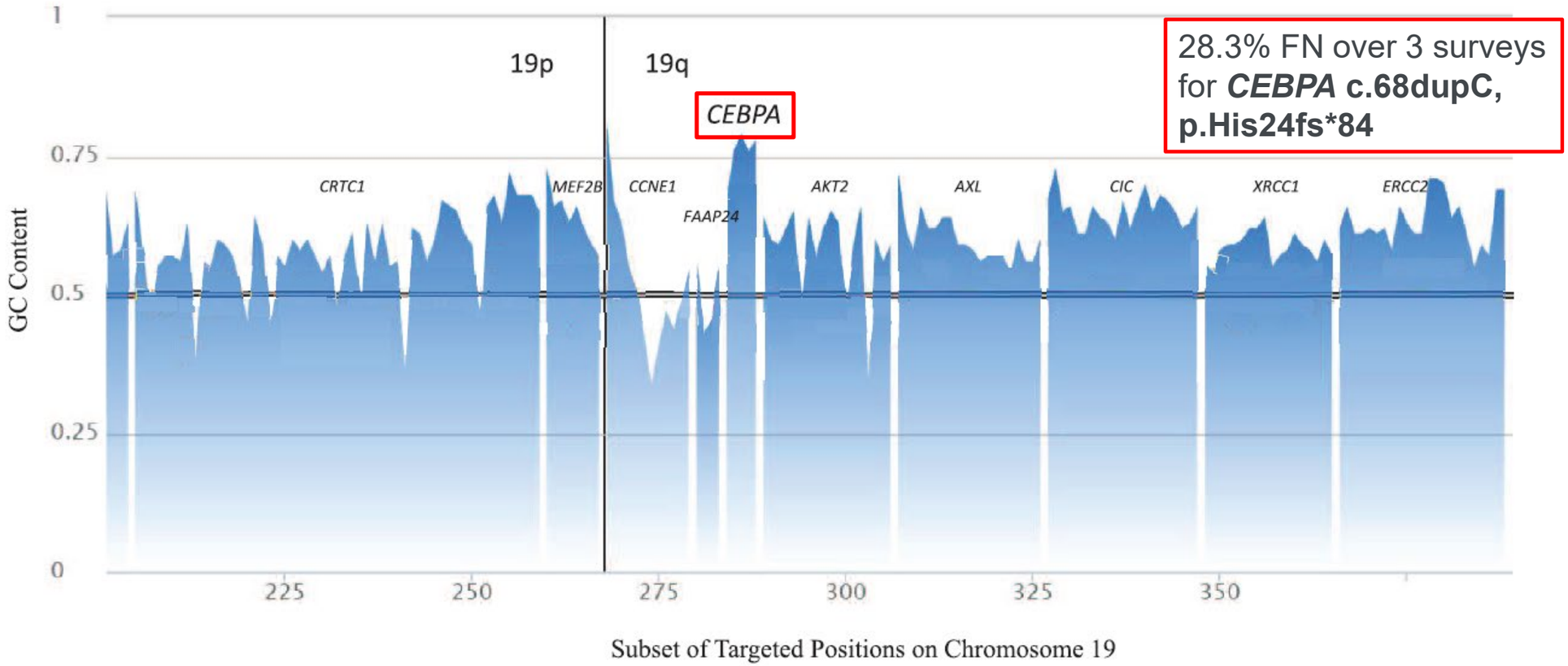
1) More thermostable (molecular interactions of base stacking), don't melt well at usual PCR denaturation T
2) Primers for GC rich regions tend to form hairpins and dimers



Source 2

2. Analytical Errors:

A. False Negative (FN) errors – Variants in difficult-to-sequence regions with high GC content



Reference: Nardi, V. *Arch Pathol Lab Med* (2022)

High GC Region “Remedies”

- ✓ Adding DMSO 2.5%-5%
- ✓ Adding 1.0M betaine
- ✓ Adding a heat-denaturation step
- ✓ Consider using an orthogonal method

Reference: Nardi, V. *Arch Pathol Lab Med* (2022)

2. Analytical Errors:

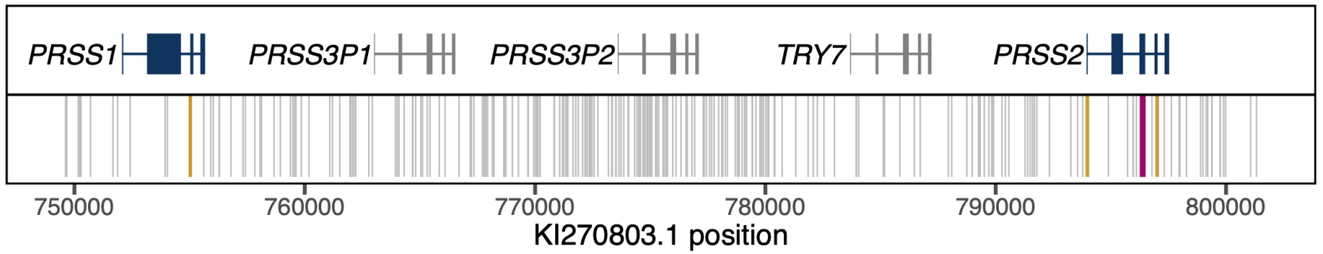
A. False Negative (FN) errors – Pseudogene interference

Pseudogenes: genomic sequences that are similar to a gene but are considered to be nonfunctional.

Owing to their sequence similarity to functional genes, pseudogenes can interfere with short-read NGS technology, resulting in mismapping of reads between the gene and pseudogene that can lead to either false negative or false-positive calls

NGS-Germline 2019-A survey included genomic position chr 7:g.142460335 (NM_002769.4), located in **PRSS1**

PRSS1 encodes a trypsinogen and has 2 known pseudogenes, PRSS3P1 and PRSS3P2



https://github.com/Shuhua-Group/NGS.PRSS1-2caller

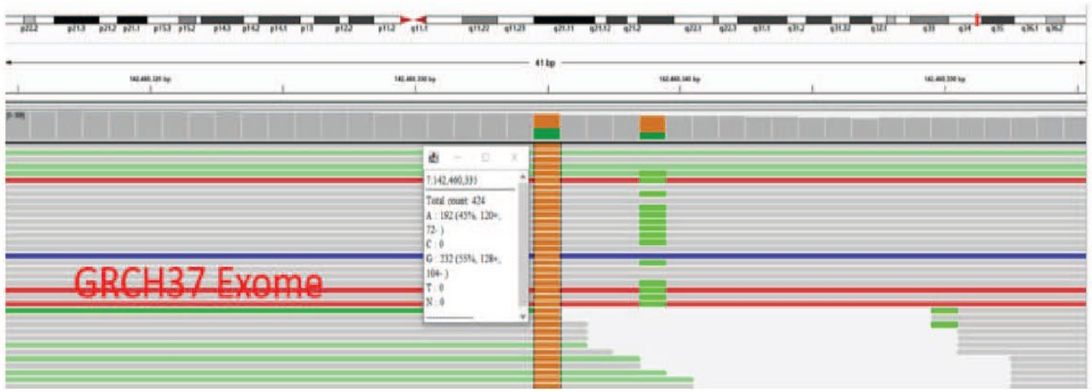
Reference: Nardi, V. *Arch Pathol Lab Med* (2022)

2. Analytical Errors:

A. False Negative (FN) errors – Pseudogene interference

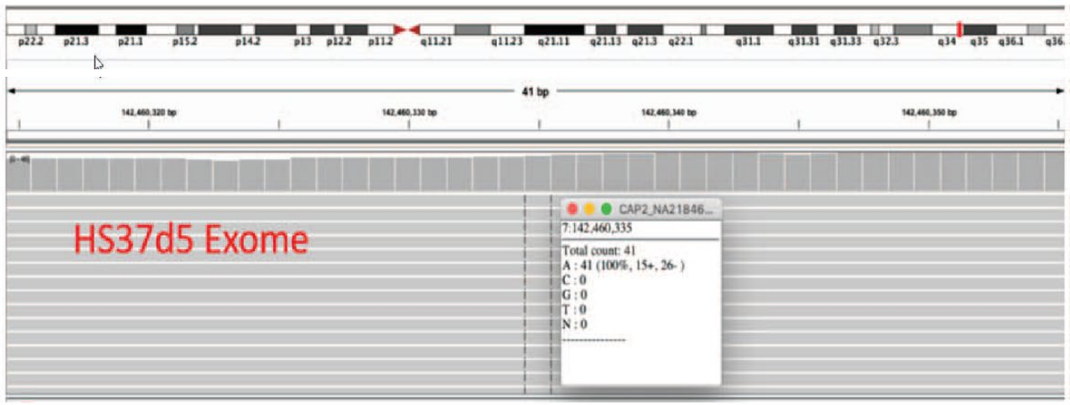
NGS-Germline 2019-A survey included genomic position chr 7:g.142460335 (NM_002769.4), located in **PRSS1**

A



51.6% (48), FP

B



47.3% (44), TN

Reference: Nardi, V. *Arch Pathol Lab Med* (2022)

Pseudogene Interference “Remedies”

- ✓ Align to the suggested/more recent reference genome
- ✓ Identify region of homology requiring specific attention
- ✓ Consider long range PCR and Sanger sequencing

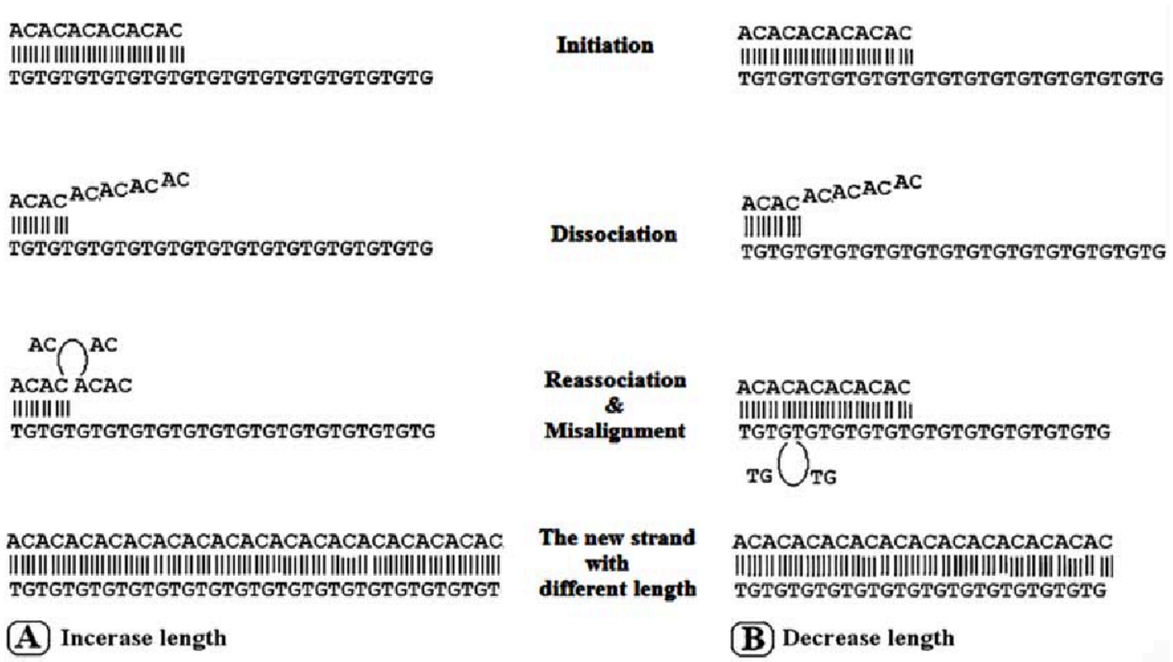
Reference: Nardi, V. *Arch Pathol Lab Med* (2022)

2. Analytical Errors:

B. False Positive (FP) errors – ie, variants in homopolymer regions

Homopolymer (HP) = mononucleotide microsatellites, a sequence of consecutive identical bases.

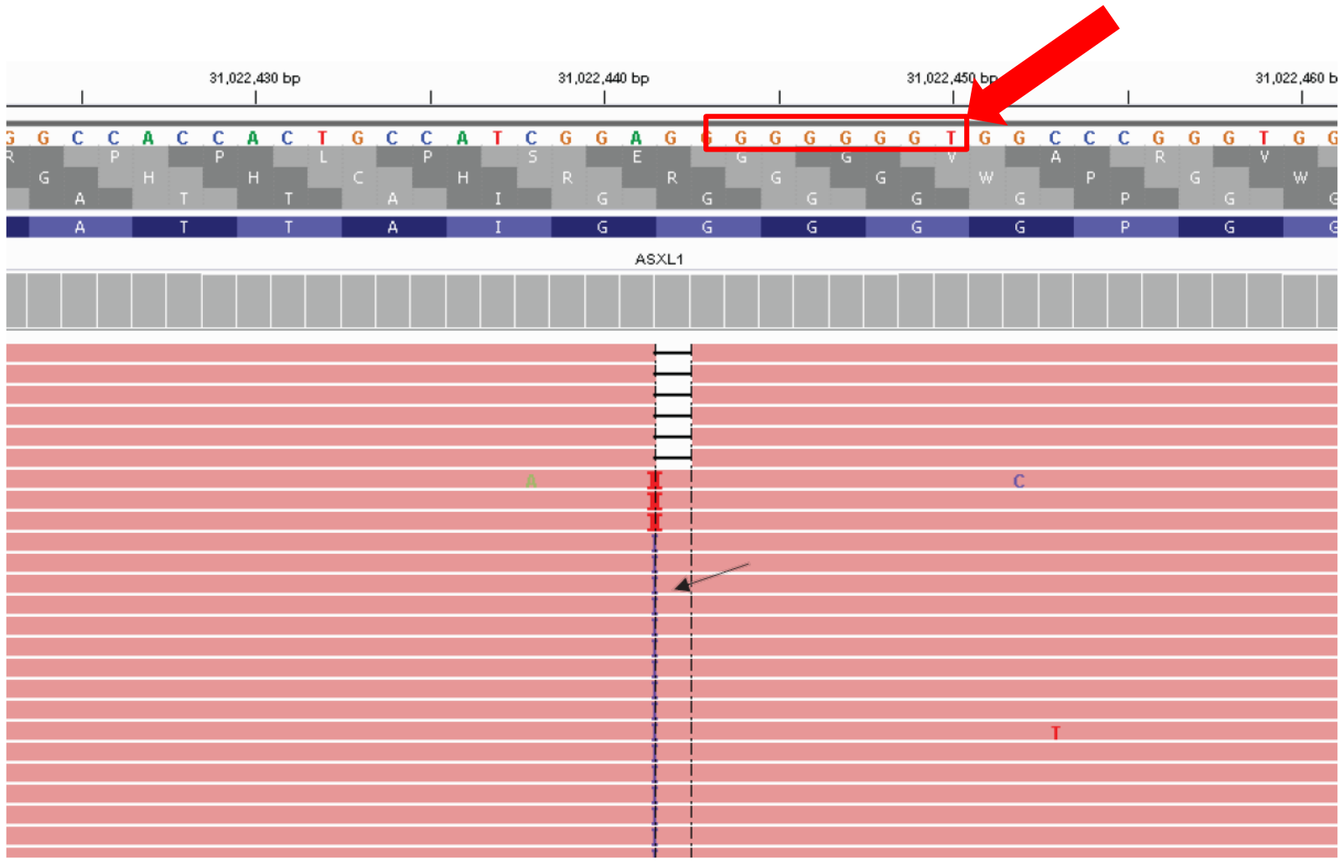
Mechanism of replication slippage from dinucleotide repeats



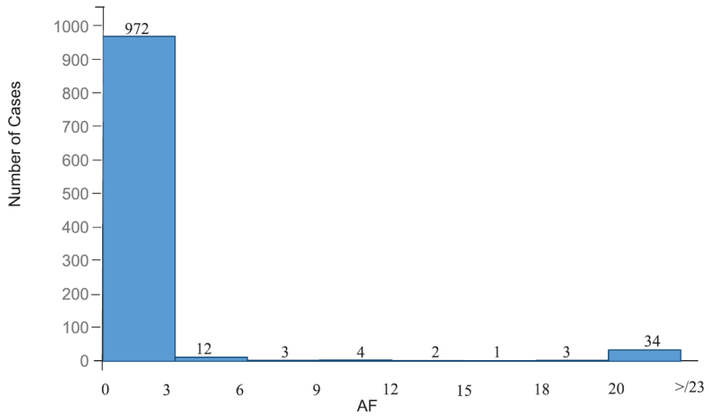
Reference: Hosseinzadeh-Colagar A, et al. *Mol Biol Res Commun.* 2016

2. Analytical Errors:

B. False Positive (FP) errors – ie, variants in homopolymer regions



ASXL1* c.1934dupG; p.Gly646fs
7.0% FP in 2018
1.3% in 2019



Reference: Nardi, V. *Arch Pathol Lab Med* (2022)

Homopolymer Region “Remedies”

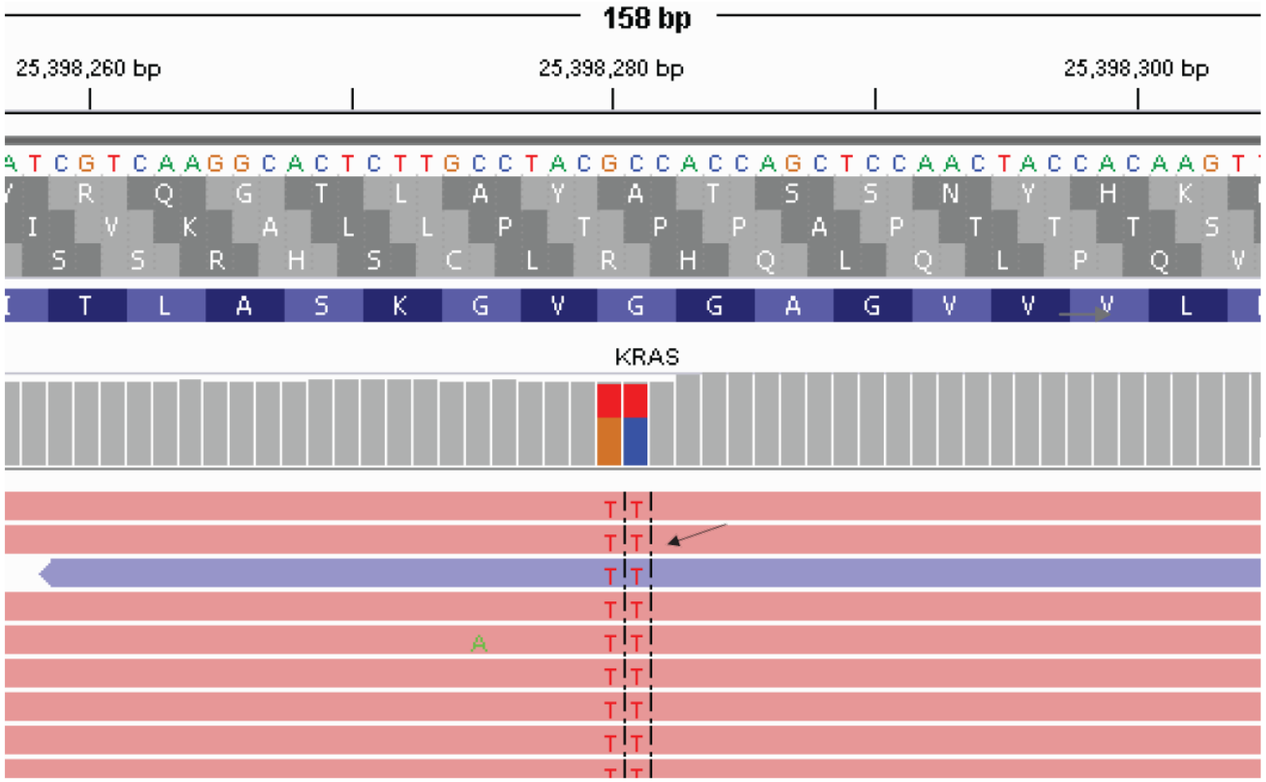
- ✓ Use high fidelity DNA polymerase
- ✓ Optimize variant calling parameters to distinguish artifacts
- ✓ Use error correction methods such as unique molecular identifiers

Reference: Nardi, V. *Arch Pathol Lab Med* (2022)

2. Analytical Errors:

C. Errors in annotation – Multinucleotide variants

A substitution changes 1 nucleotide into 1 other nucleotide; thus 2 sequential nucleotide changes (dinucleotide changes) are not considered substitutions but rather deletion-insertion (delins) variants.



Correct annotation:
***KRAS* c.38_39delinsAA, p.Gly13Glu**

11.9-37.5% of laboratories did not correctly report multinucleotide variants

Reference: Nardi, V. *Arch Pathol Lab Med* (2022)

Laboratory Performance for Detection of Dinucleotide Variants

Gene	Transcript	Nucleotide change	Protein change	Chromosomal position	No. of labs that tested for variant	No. (%) of labs that detected variant	No. (%) of labs that missed the variant	Engineered VAF, %	Average Reported VAF, %	Median Coverage	Mailing
CDKN2A	NM_000077.4	c.171_172delCCinsTT	p.Arg58*	chr9:21971186_21971187delGGinsAA	32	20 (62.5)	12 (37.5)	20.0	16.2	1997.0	NGSB1/2 A 2018
HRAS	NM_005343.2	c.37_38delGGinsAA	p.Gly13Asn	chr11:534285_534286delCCinsTT	33	25 (75.8)	8 (24.2)	25.0	20.6	1120.0	NGSB1/2 A 2019
HRAS	NM_005343.2	c.181_182delCAinsTT	p.Gln61Leu	chr11:533874_533875delTGinsAA	32	28 (87.5)	4 (12.5)	30.0	26.8	1996.0	NGSB1/2 B 2018
KRAS	NM_004985.3	c.38_39delGCinsAA	p.Gly13Glu	chr12:25398280_25398281delGCinsTT	111	85 (76.6)	26 (23.4)	14.2	13.7	1985.0	NGSHM A 2019
KRAS	NM_004985.3	c.180_181delTCinsAA	p.Gln61Lys	chr12:25380277_25380278delGAinsTT	42	37 (88.1)	5 (11.9)	10.0	8.5	2955.0	NGSB1/2 A 2018
NRAS	NM_002524.4	c.182_183delTTinsTG	p.Gln61Pro	chr1:115256528_115256529delAAinsTG	31	26 (83.9)	5 (16.1)	45.0	43.1	6888.6	NGSB1/2 A 2016
NRAS	NM_002524.4	c.180_181delACinsTA	p.Gln61Lys	chr1:115256530_115256531delGTinsTA	39	31 (79.5)	8 (20.5)	20.0	16.8	1987.0	NGSB1/2 A 2019

Reference: Nardi, V. *Arch Pathol Lab Med* (2022)

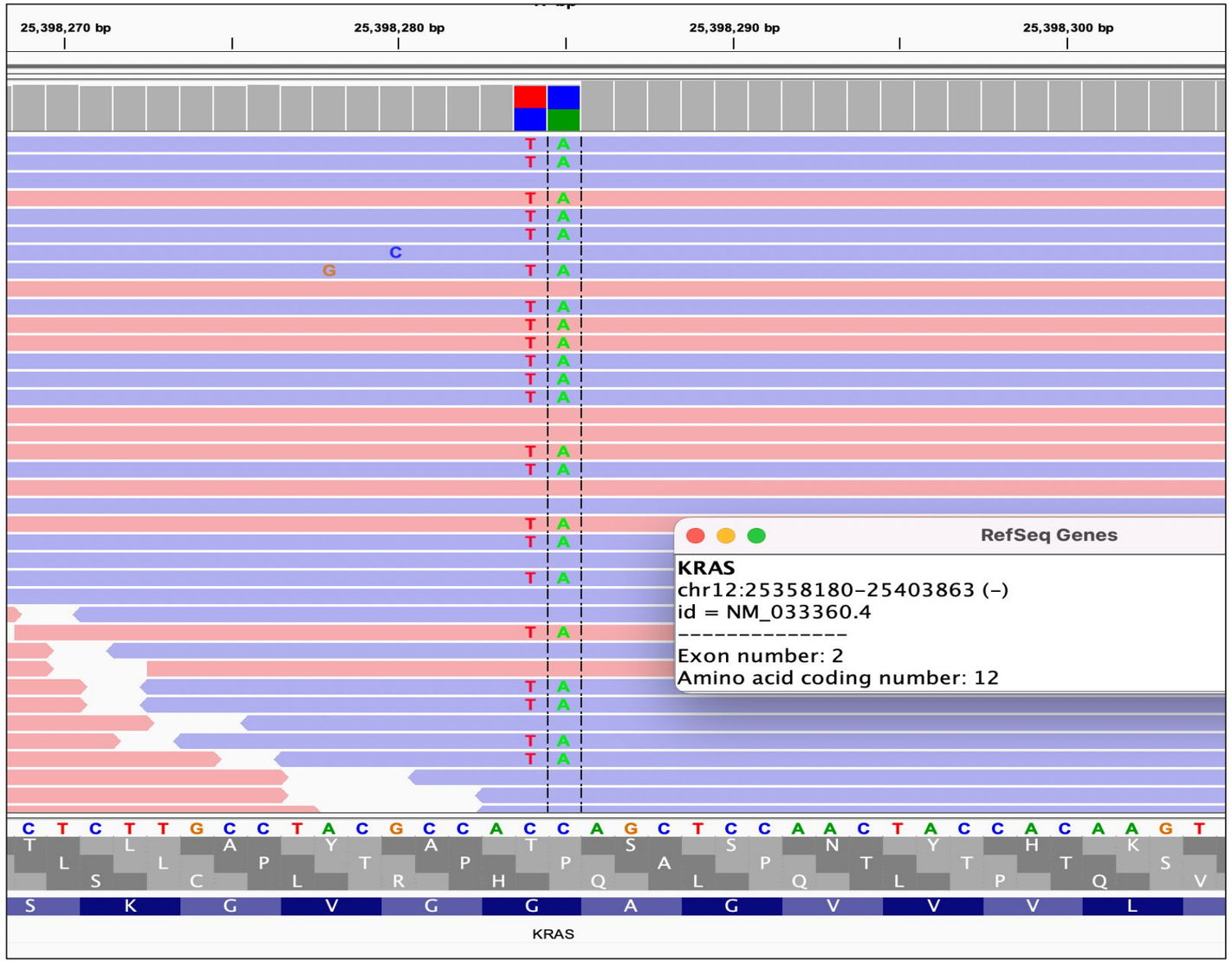
Does the Patient have a targetable *KRAS* p.Gly12Cys mutation? *Which one is the Correct Annotation?*

A. KRAS c.34C>A p.Gly12Cys & KRAS c.35C>T, p.Gly12Asp

B. KRAS c.34G>T p.Gly12Cys & KRAS c.35G>A, p.Gly12Asp

C. KRAS c.34_35delGGinsTA, p. Gly12Tyr

D. KRAS c.34_35GG>TA, p. Gly12Tyr

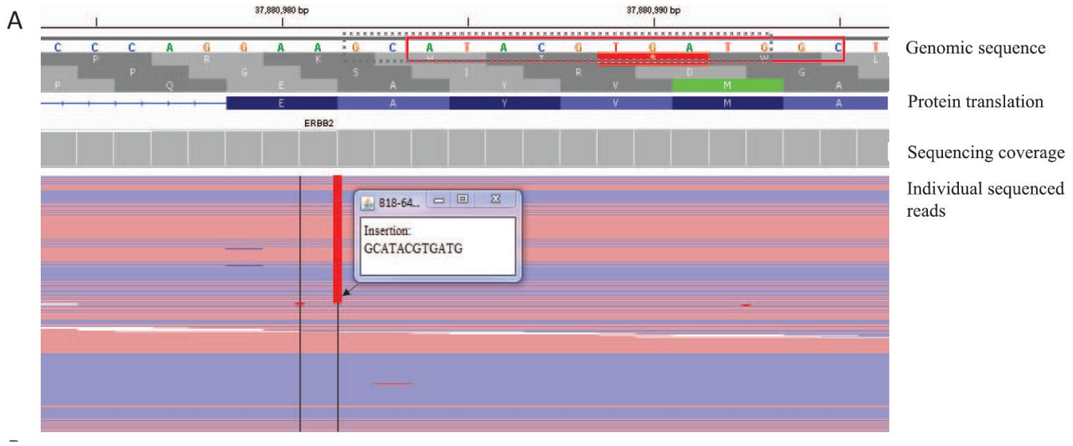


2. Analytical Errors:

C. Errors in annotation – Duplicating insertions

Insertions that duplicate the immediately preceding nucleotide or sequence should be described as **duplications**, not as insertions (per HGVS).

3' rule: the most 3' position possible is arbitrarily assigned to be where the duplication occurs, important when the duplication involves stretches of tandem repeats.



ERBB2 duplication
(NM_004448.2:c.2313_2324dup ATACGTGATGGC;
p.Tyr772_A775dupTyrValMetAla)
25.8% error rate in 2018 (25.0% VAF)
12.4% in 2019 (39% VAF)

GAAGCATACGTGATGGCT

GAAGCATACGTGGCATACGTGATGGCT Insertion

GAAGCATACGTGGCATACGTGATGGCT 3' rule

Reference: Nardi, V. *Arch Pathol Lab Med* (2022)

Multinucleotide Variants and Duplication Variants' "Remedies"

- ✓ Manual review of variants
- ✓ Appropriate use of current HGVS nomenclature

Reference: Nardi, V. *Arch Pathol Lab Med* (2022)

3. Post-analytic Errors:

Transcription errors (Though some of these errors could also be pre- analytic and due to specimen swap)

0–2.0% per Survey

Table 5. Percentage of Proficiency Testing Participants With Specimen Swaps and/or Transcription Errors by Survey and Series

	Mailing							
Proficiency Test	2016-A	2016-B	2017-A	2017-B	2018-A	2018-B	2019-A	2019-B
NGSHM	0.0% (0/57)	0.0% (0/59)	1.2% (1/81)	0.0% (0/87)	2.0% (2/99)	0.0 % (0/101)	1.6% (2/122)	1.4% (2/141)
NGSST	0.0% (0/116)	0.0% (0/120)	0.6% (1/154)	1.2% (2/171)	0.0% (0/188)	0.5% (1/197)	1.0% (2/195)	0.0% (0/205)

Abbreviations: NGSHM, next-generation sequencing hematologic malignancies; NGSST, next-generation sequencing solid tumor.
Numbers are presented in parentheses.

Reference: Nardi, V. *Arch Pathol Lab Med* (2022)

Transcription Errors Don't Occur Just on Surveys

"Incidents in Molecular Pathology"
(2018 European cross-sectional study of 8 labs)

Table 3. Causes of Incidents Reported for NSCLC and mCRC Biomarker Testing

Incident Cause (N = 822)	n	%
Preanalytical	166	20.2
Samples switched	35	21.1
Incorrect or missing sample labels during cutting or DNA extraction	31	18.7
Sample microtomy problems (eg, incorrect protocol/tissue used, floaters, cut too deep, too thick, technical issues, etc)	29	17.5
Problem during sample embedding (eg, contamination, damage, multiple tissues in one block, etc)	17	10.2
Lost material	16	9.6
Equipment problems (errors, defects, dispensing of reagents, etc)	14	8.4
Accompanying H&E stain missing, bad quality, or incorrect	9	5.4
Inadequate amount of material for cutting/ extraction	8	4.8
Sample delayed between departments/ services	7	4.2

Table 3. Causes of Incidents Reported for NSCLC and mCRC Biomarker Testing

Incident Cause (N = 822)	n	%
Analytical	275	33.5
Failed IHC or FISH test	42	15.3
Technical/server problems with autostainer or sequencer	44	16.0
Missing, inadequate, or expired reagents	39	14.2
Incorrect sample labelling/worksheet	28	10.2
Sample switch	17	6.2
Sample lost or not tested	16	5.8
Problem with procedure (unspecified)	11	4.0
Inadequate, failed, or lack of control tissue	21	7.6
Faint/too much background FISH/IHC signal	18	6.5
Incorrect test performed/procedure not followed	11	4.0
Insufficient/inadequate material to perform the test	9	3.3
Failed or incorrect sequencing run	7	2.5
Sample contamination	7	2.5
Other	5	1.8

Table 3. Causes of Incidents Reported for NSCLC and mCRC Biomarker Testing

Incident Cause (N = 822)	n	%
Postanalytical	194	23.6
Report content	152	78.4
Incorrect result/conclusion on report	44	28.9
Patient information (name or date of birth) incorrect	27	17.8
Incorrect validation (too soon or too late) of report	18	11.8
Incorrect sample localization	15	9.9
Error in microscopy part (unspecified)	13	8.6
Incorrect or absent sample number on report, but correct result	13	8.6
Incorrect clinical history	7	4.6
Missing result on report	6	3.9
Incorrect report template used	3	2.0
Incorrect requesting physician mentioned	3	2.0
Incorrect author of report	3	2.0
Reports/results from patients switched	16	8.2
Software problem with automated result	13	6.7
No report present	6	3.1
Unexplained molecular result obtained	5	2.6
Documented procedure on reporting lacking	2	1.0

Reference: Keppens, C. *Arch Pathol Lab Med* (2021)

Post-analytic/Transcription Errors' "Remedies"

- ✓ Have a second person check every entry before submission

Reference: Nardi, V. *Arch Pathol Lab Med* (2022)

Summary of the Most Common Challenges and Remedies

Table 6. Recurrent Proficiency Testing Challenges and Possible Remedies

Challenge	Remedy
Detection of variants in genomic region with high GC content, difficult to sequence	Add 2.5%–5.0% DMSO Add 1.0 M betain Add a heat-denaturation step Consider use of an orthogonal method
Detection of variants from homopolymer regions	Use of a high-fidelity DNA polymerase Optimization of variant calling parameters to distinguish artifacts Use of error correction methods, such as unique molecular identifiers
Missing variants or false positives from pseudogene interference	Align to the hs37d5 reference genome Identify region of homology requiring specific attention Consider long-range PCR and Sanger sequencing
Errors in reporting dinucleotide variants	Manual review of variants Appropriate use of current HGVS nomenclature
Errors in reporting duplication variants as insertions	Manual or bioinformatic review of the raw data Appropriate use of current HGVS nomenclature
Errors due to use of different transcript	Report the transcript and version used
Postanalytic errors due to specimen swaps or transcription errors	Conduct a critical analysis of potential steps that could lead to nonanalytic errors Have a second person check every entry before submission Avoid multiple patient specimens in the active work area at the same time Label only 1 specimen at a time before proceeding to the next specimen Have a second person check the labeling of tubes Consider investigating potential sample swaps with molecular methods

Abbreviations: DMSO, dimethyl sulfoxide; HGVS, Human Genome Variation Society; PCR, polymerase chain reaction.

Reference: Nardi, V. *Arch Pathol Lab Med* (2022)

Summary

- Errors in clinical laboratories performing NGS are overall uncommon!!
- Most could be avoided by:
  - Optimizing specimen adequacy assessment
  - Optimizing informatic pipelines
  - Manual review of results
  - Using the correct mutation nomenclature

COLLEGE of AMERICAN PATHOLOGISTS

Thank you. Contact us!

international@cap.org

+1-847-832-7000

References

- Valentina Nardi, Karen D. Tsuchiya, Annette S. Kim, Lora J. H. Bean, Jaimie G. Halley, Thomas A. Long, Szabolcs Szelinger, Patricia Vasalos, John A. Thorson, Ann M. Moyer, Joel T. Moncur; Next-Generation Sequencing Somatic and Germline Assay Troubleshooting Guide Derived From Proficiency Testing Data. *Arch Pathol Lab Med* 1 April 2022; 146 (4): 451–461. doi: https://doi.org/10.5858/arpa.2020-0842-CP.
- Alissa Keegan, Julia A. Bridge, Neal I. Lindeman, Thomas A. Long, Jason D. Merker, Joel T. Moncur, Nathan D. Montgomery, Rakesh Nagarajan, Paul G. Rothberg, Mark J. Routbort, Patricia Vasalos, Rena Xian, Annette S. Kim; Proficiency Testing of Standardized Samples Shows High Interlaboratory Agreement for Clinical Next-Generation Sequencing–Based Hematologic Malignancy Assays With Survey Material–Specific Differences in Variant Frequencies. *Arch Pathol Lab Med* 1 August 2020; 144 (8): 959–966. doi: https://doi.org/10.5858/arpa.2019-0352-CP.
- Moore RA, Zeng T, Docking TR, et al. Sample Tracking Using Unique Sequence Controls. *J Mol Diagn*. 2020;22(2):141-146. doi:10.1016/j.jmoldx.2019.10.011
- Kelly A. Devereaux, Rhona J. Souers, Rondell P. Graham, Bryce P Portier, Lea F. Surrey, Anna Yemelyanova, Patricia Vasalos, Dimitri G. Trembath, Joel T. Moncur; Neoplastic Cellularity Assessment in Molecular Testing: A Multi-institutional Practice Survey and Performance Challenge Identifies a Need for Standardization. *Arch Pathol Lab Med* 1 September 2022; 146 (9): 1062–1071. doi: https://doi.org/10.5858/arpa.2021-0166-CP.
- Hosseinzadeh-Colagar A, Haghighatnia MJ, Amiri Z, Mohadjerani M, Tafrihi M. Microsatellite (SSR) amplification by PCR usually led to polymorphic bands: Evidence which shows replication slippage occurs in extend or nascent DNA strands. Mol Biol Res Commun. 2016;5(3):167-174.
- Keppens, C., Van Royen, Y., Brysse, A., Cotteret, S., Høgdall, E., Kuhlmann, T. P., O'Sullivan, B., Pauwels, P., Pauwels, S., Rot, M., Vanderheyden, N., Van Hee, I., & Dequeker, E. M. (2021). Incidents in Molecular Pathology: Frequency and Causes During Routine Testing. *Arch Pathol Lab Med 2021 Oct 1*; 145(10), 1270–1279. https://doi.org/10.5858/arpa.2020-0152-OA.

Sources

1. Javed N et al., Nature Communications 2020
2. https://en.wikipedia.org/wiki/GC-content#/media/File:AT-GC.jpg