Improved sensitivity for clinically relevant variants using orthogonal sequencing

John F. Thompson[1], Niru Chennagiri[1], Eric J. White[1], Edgardo Lopez[1], Anastasia Nikiforov[1], Tristen Ross[1], Rebecca Batorsky[1], Laura Fahy[1], Sherry Hansen[1], Seyed Ali Hosseini[1,3], Va Lip[1], David Margulies[1,2,3], Nichole Napolitano[1], Marcia M. Nizzari[1], Joe Strycharz[1], Natalie Vena[1], Timothy Yu[1]

[1]Claritas Genomics, Cambridge MA; [2]Harvard Medical School, Boston MA; [3]Boston Children's Hospital, Boston MA

CLARITAS GENOMICS

ABSTRACT

We have previously shown that using independent Next Generation Sequencing platforms provided improved sensitivity for detecting exomic variants in the well-defined regions of standard reference samples. However, it is not straightforward to compare results in other parts of the exome and extrapolate to clinically relevant variants. To assess the ability of dual platforms to improve results, we examined the likely/possibly pathogenic variants found in more than 100 exome-based Regions of Interest. As previously observed, ~95% of all variants were orthogonally confirmed with identical sequencing results found on both the Illumina NextSeq and the Ion Torrent Proton platforms.

Variants reported to patients included those which were sufficiently well studied as to be classified as Pathogenic/Likely Pathogenic by ACMG guidelines. In addition, Variants of Uncertain Significance (VUS) were found in genes known to be disease-causing but the specific variants could not be unambiguously classified as pathogenic. These variants were also reported. Among the P/LP/VUSs, 96% were orthogonally confirmed but there were also examples of variants found exclusively on the NextSeq (3.2%) or exclusively on the Proton (0.8%). **If only a single platform had been used, these variants would have been missed and not reported to the patient.** The number of genes that are 100% covered at >20x increased by 42% when the Proton coverage was added to the Illumina analysis. Thus, this method is shown to deliver confirmed variants in a timely fashion while expanding the sensitivity for identifying rare, disease-relevant variants.

Orthogonally-confirmed Variants Approach 100% PPV

RefSeq intersection NIST intersection CRE intersection AmpliSeq

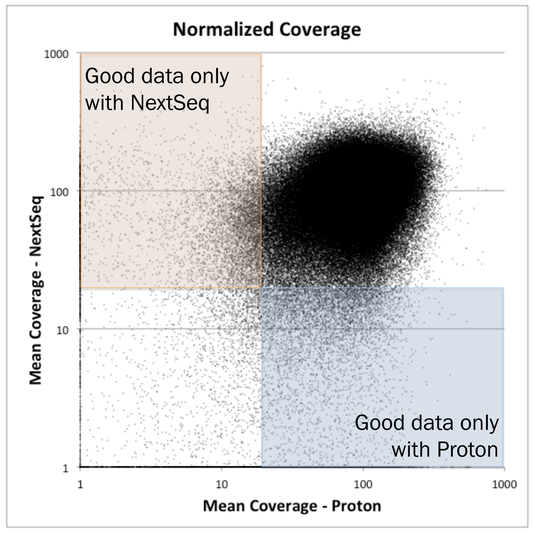
Category	% of Total	# FP	# FPs Fixed by Sanger	# TP	PPV
Orthogonally Confirmed – Proton call matches NextSeq Pass call	94.4%	1	0	49167	99.998%
Reliable - Proton call matches NextSeq NoPass	0.3%	0	0	134	97.81%
Likely True Positives - Singleton NextSeq call or Singleton Proton call with no NextSeq coverage	4.6%	124	ND	2249	94.77%
Likely False Positives - Singleton NextSeq NoPass or Singleton Proton call with NextSeq coverage	0.8%	346	ND	79	18.59%

Platform-specific Sensitivity

	RefSeq ∩ NIST v2.19 ∩ CRE ∩ AmpliSeq						
Type	**SENS**	**SPEC (FP/MB)**	**PPV**	**# FPs**	**# TPs**	**# TNs**	**# FNs**
	Illumina MiSeq						
SNV	98.99%	1.75	99.71%	48	16585	27431566	168
InDel	92.79%	0.53	96.90%	15	459	27431566	36
All	98.82%	2.29	99.63%	63	17044	27431566	204
	Illumina NextSeq						
SNV	99.60%	1.85	99.70%	51	16704	27431546	67
InDel	95.00%	0.55	96.90%	15	469	27431546	25
All	99.47%	2.39	99.62%	66	17173	27431546	91
	Proton Filtered						
SNV	96.89%	2.66	99.55%	73	16251	27431518	521
InDel	50.96%	0.78	92.20%	21	254	27431518	244
All	95.61%	3.44	99.43%	94	16505	27431518	759
	NextSeq/Proton Combined						
SNV	99.88%	4.31	99.30%	118	16753	27431457	20
InDel	95.01%	1.32	92.83%	36	469	27431457	25
All	99.74%	5.64	99.11%	155	17222	27431457	45

When coverage differences between the MiSeq (46x) and NextSeq (125x) are factored in, the two instruments have nearly identical performance (sensitivity and specificity). Proton (133x) specificity is similar to the Illumina instruments but sensitivity is lower for InDels. The Proton results include a custom filter that eliminates common instrument-specific artifacts. When the NextSeq and Proton platforms are combined, SNV sensitivity is 99.88% for the NIST consensus region. InDel sensitivity is ~95%.

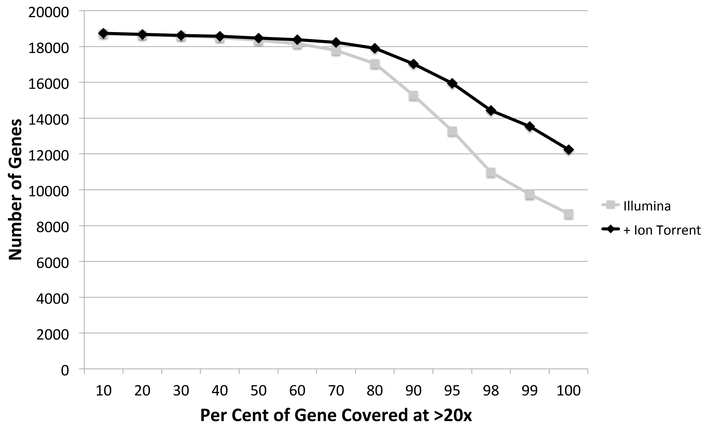
Platform-Specific Coverage



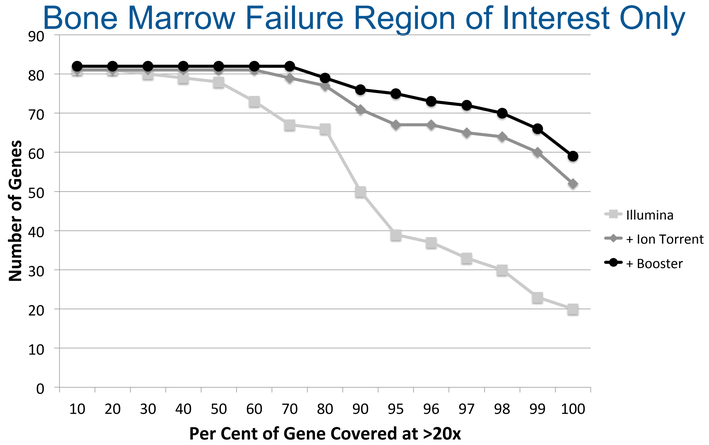
~90% of 187,475 exons have >20x coverage on both platforms when normalized to 100x coverage on each. ~2% of exons are covered <20x on both platforms. ~7% exons are covered >20x on one platform but not the other. Thus, sensitivity increases 3-4% when both platforms are used together compared to either alone. The singly covered regions allow a greater percentage of variants to be identified and confirmed by other methods.

Coverage Improvements with Added NGS Platform

Whole Exome



Bone Marrow Failure Region of Interest Only



Ideally, all genes of interest would be covered 100% at >20x but many genes typically have short segments that are not well covered. When a second NGS platform is used to supplement the Illumina NGS, the number of genes 100% covered across the whole exome increases by 42%. In the more defined (and challenging) BMF ROI, well over twice as many genes are covered 100%. Furthermore, the ROI can be boosted with custom amplicons to increase coverage even further.

Rare, Clinically-relevant Variants Confirm at Same Frequency as Common Variants

Platform with Correct Variant Call	Orthogonally Confirmed - Pathogenic/Likely Path	Orthogonally Confirmed - VUS	No/Poor Coverage on Other Platform	Reference Call on Other Platform		Zygosity/ Call Mismatch
				SNP	InDel	
Illumina	45	432	8	0	4	1
Ion Torrent			3	0	0	

Variants reported for 77 patient DNAs analyzed using exome-based neurological Regions of Interest were examined for how the variants were identified. The minimum number of genes examined in these samples was 196 and, depending on the phenotypes chosen, typically included 250-600 genes for each DNA.

96% of the Pathogenic/Likely Pathogenic/VUSs were identified on both Illumina and Ion Torrent NGS platforms so did not require additional confirmation. This frequency is very similar to the 94% seen with all variants in the NIST region of NA12878. Most of the variants found on just one NGS platform were not covered on the other platform though the Ion Torrent also missed some InDels that were in the covered region. Similar results were found with other phenotypically-defined Regions of Interest. The singleton variant calls were all confirmed via Sanger sequencing prior to reporting.

SUMMARY

The benefits of broad genetic testing of patients for previously undiagnosed diseases for the patient are clear. However, the quality and costs of such testing and reimbursement for them have been problematic. Whole Exome Sequencing (WES) provides cost-efficient identification of clinically-relevant variants that eliminates the high expense of testing only a few genes at a time and the resultant lengthy diagnostic odysseys. Orthogonal DNA selection and sequencing methodologies provides better sensitivity than standard WES as well as immediate confirmation of ~95% of all variants. This improves turnaround time and eliminates the need and cost for most subsequent Sanger confirmation.

While Sanger sequencing is generally considered the gold standard for confirmation, we (and others) have found that it is subject to the same amplification and repeat-based artifacts that can afflict NGS technologies. These difficulties highlight the issue that even the "gold standard" Sanger sequencing technology is error-prone and subject to artifacts. **Thus, the use of a truly independent NGS platform to confirm putative variants is preferable to Sanger sequencing for reasons of cost and speed.**

In addition to the immediate confirmation of ~95% variants and the high accuracy of orthogonally-confirmed variants, another key benefit to the parallel exome sequencing is the increased sensitivity due to the overlapping regions that are covered by each platform. Because the NIST reference is biased for regions that are most easily sequenced, results can be deceptive and overestimate the true sensitivity of both platforms. The singly covered regions allow a greater percentage of variants to be identified and subsequently confirmed by other methods. **There are thousands of variants in the exome that are detected only by the NextSeq or only by the Proton.** These variants require confirmation prior to clinical reporting but they would not have been reported at all if only the NextSeq or only the Proton had been used exclusively. By using orthogonal and complementary technologies, we are able to quickly confirm variants at a genome-wide scale and provide improved sensitivity for detecting potentially pathogenic variants.

Feel free to:

CLARITAS GENOMICS INC.
99 Erie Street | Cambridge, MA 02139

clientservices@claritasgenomics.com
p: 617.553.5880 p: 855.373.9003 f: 617.553.5842