The ACE Clinical Exome: An Augmented Exome Providing Accurate Structural Variant Detection

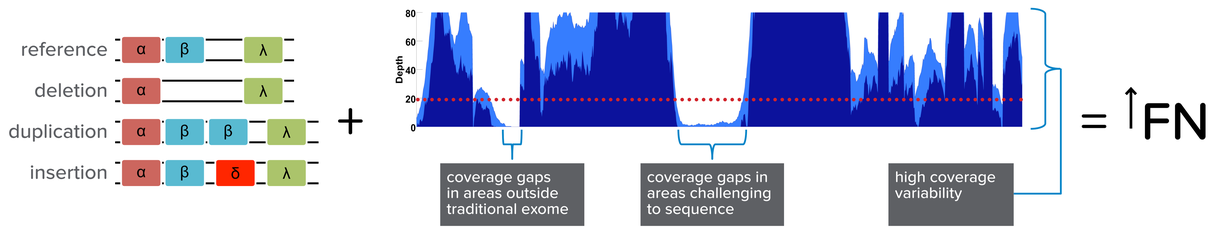
Anil Patwardhan, Stephen Chervitz, Ming Li, Jason Harris, Gabor Bartha, Daniel Newburger, Mark Pratt, Sarah Garcia, Jeanie Tirch, Nan Leng, Christian Haudenschild, Shujun Luo, Deanna M. Church, John West and Richard Chen

Personalis, Inc. | 1350 Willow Road, Suite 202 | Menlo Park, CA 94025

Contact: **apatwardhan@personalis.com**

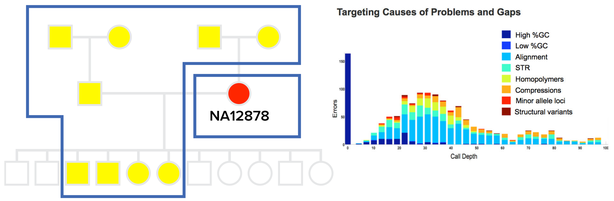
Introduction

Whole-Exome Sequencing (WES) has become a valuable tool in identifying common and rare disease-causing variants due to its broad coverage and high-resolution. However, copy number variations (CNVs), a form of structural variation (SV) that leads to abnormal copies of genomic regions, are not reliably detected using standard WES approaches. Limitations of standard WES lead to high **false-negative rates.**

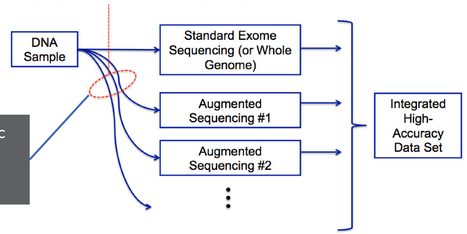


Methods in Constructing and Evaluating the ACE Augmented Exome

Enhance coverage in regions difficult to sequence

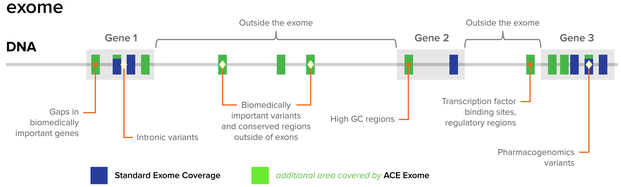


Augmented Exome Approach optimizes fixing systematic issues like GC bias by doing capture under different sample prep conditions

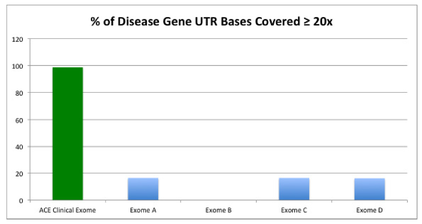


target problematic regions/gaps in medically associated genes

Extend coverage to medically relevant areas outside the traditional exome

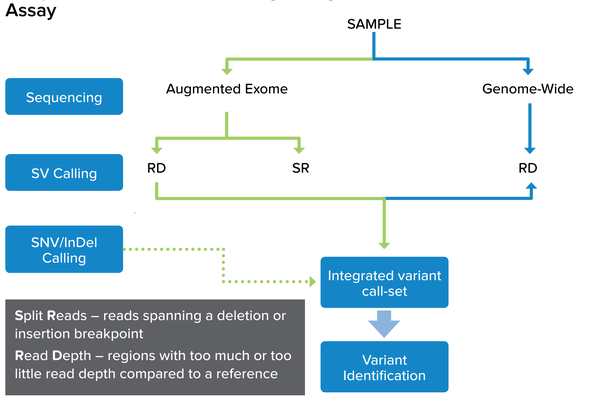


targets include CNV "hotspots" (SV junction sequences, segmental duplications, disease associated intergenic/intronic regions) and UTRs



ACE Augmented Exome covers more disease gene UTR regions than other exomes

Integrate CNV Detections Using a Targeted Exome and Genome-wide Assay

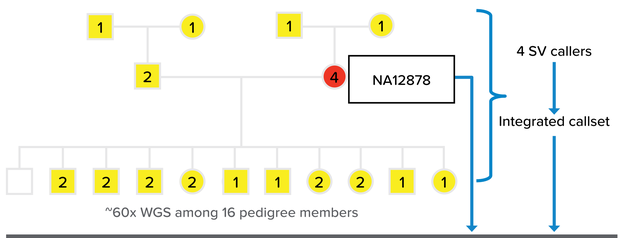


Split **R**eads – reads spanning a deletion or insertion breakpoint

Read **D**epth – regions with too much or too little read depth compared to a reference

Creating an SV-gold Set for Evaluation

NIST GIBv2.181 is the recognized genomic scale standard, but is limited by variant type and regions covered. There is no widely-accepted "high-confident" call-set for SVs



2345 Deletions

in NA12878 and 2 other samples with exact breakpoint matches vetted by inheritance state and variant type consistency

[1]Zook *et al.*, Integrating human sequence data sets provides a resource of benchmark SNP and indel genotype calls. Nat. Biotech. 32: 246–251 (2014)

Results

Known Pathogenic Variants

Evaluation of the ACE Augmented Exome Using a Set of 28 Reference Samples Harboring 30 CNVs

SAMPLE, PHENOTYPE	LOCATION	SIZE	TYPE	RD	SR	RD
GM13019; Turner Syndrome	chrX:62260103-153703648	$9.1x10^7$	DEL	N	N	Y
GM20022; Multiple Clinical Features	chr3:136044785-197137370	$6.1x10^7$	DUP	Y	N	Y
GM05067; Trisomy 9	chr9:36587-44806024	$4.5x10^7$	DUP	N	N	Y
GM12606; Partial trisomy 13	chr13:17943628-59139422	$4.1x10^7$	DUP	Y	N	Y
GM09367; Multiple Clinical Features	chr6:107861056-143105847	$3.5x10^7$	DUP	Y	N	Y
GM13783; Down Syndrome	chr21:13286389-46887579	$3.4x10^7$	DEL	N	N	Y
GM01201; Down Syndrome phenotype	chr21:13322592-46921373	$3.4x10^7$	DEL	N	N	Y
GM14485; Multiple congenital anomalies	chr8:12572787-43719525	$3.1x10^7$	DEL	Y	N	Y
	chr8:160290-7213701	$7.1x10^6$	DEL	Y	N	N
GM11419; Tetrosomy Y	chrY:2712722-27209311	$2.4x10^7$	DEL	N	N	Y
	chr4:145040166-145270061	2.3x105	DEL	Y	N	N
GM22601; Wolf-Hirschhorn	chr4:55665-25591051	$2.6x10^7$	DEL	Y	N	Y
GM10925; Greig Cephalopolyssyndactyly	chr7:38598541-54681998	$1.6x10^7$	DEL	Y	N	Y
GM16595; Cri-Du-Chat	chr5:8686804-24072399	$1.5x10^7$	DEL	Y	N	Y
GM09888; Trichorhinophalangeal	chr8:107189214-119363784	$1.2x10^7$	DEL	Y	N	Y
GM07945; Adenosine Deaminase Def.	chr20:32961915-44293878	$1.1x10^7$	DEL	Y	N	Y
GM21698; Multiple congenital anomalies	chr6:162860228-170761408	$7.9x10^6$	DEL	Y	N	Y
GM21887; Angelman Syndrome	chr15:20224751-26500067	6.3x106	DEL	Y	N	Y
GM22624; Potocki-Shaffer	chr11:40433344-46031324	5.6x106	DEL	Y	N	Y
GM22991; Chromosome 1p36 Deletion	chr1:742429-5215341	4.5x106	DEL	Y	N	Y
GM13476; Smith-Magenis	chr17:16704280-20336467	$3.6x10^6$	DEL	Y	N	Y
GM12662; Seizures; mental retardation	chrX:151659961-154582680	2.9x106	DUP	Y	N	Y
GM17942; DiGeorge Syndrome	chr22:17030682-19792611	2.8x106	DEL	Y	N	Y
GM13464; Williams-Beuren	chr7:72363697-73780028	$1.4x10^6$	DEL	Y	N	Y
NA13553; Prader-Willi	15q12	10^5	DEL	Y	Y	Y
NA01712; Cockayne	10q11; del665_723	10^4	DEL	N	N	Y
NA17819; Adrenoleukodystrophy	Xq28; exons 8-10	10^4	DEL	Y	N	N
NA20212; Langer Mesomelic Dysplasia	Xpter-p22.32	10^4	DEL	Y	N	N
NA20381; Ceroid Lipofuscinosis	16p12.1	10^4	DEL	Y	N	N
NA12217; Adrenal Hyperplasia	6p21.3; exon 3 and 4	10^3	DEL	N	Y	Y

targeted exome and genome-wide assays collectively detect all CNVs among these cases

larger deletions/duplications like anueploidies are not detected by exome based RD methods but are captured by the genome-wide assay

smaller deletions, below the level of resolution of the rapid genome-wide assay are captured with the targeted exome

genome-wide

targeted exome

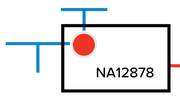
Simulation

Evaluation of the targeted exome using a simulated[2] truth set of 66 deletions ranging from 1kB – 1mB

50% reciprocal overlap	sensitivity	FDR
Targeted Exome	64%	36%

SV-Gold Set

Evaluation of the ACE Augmented Exome Using the Internally Developed SV Gold-set

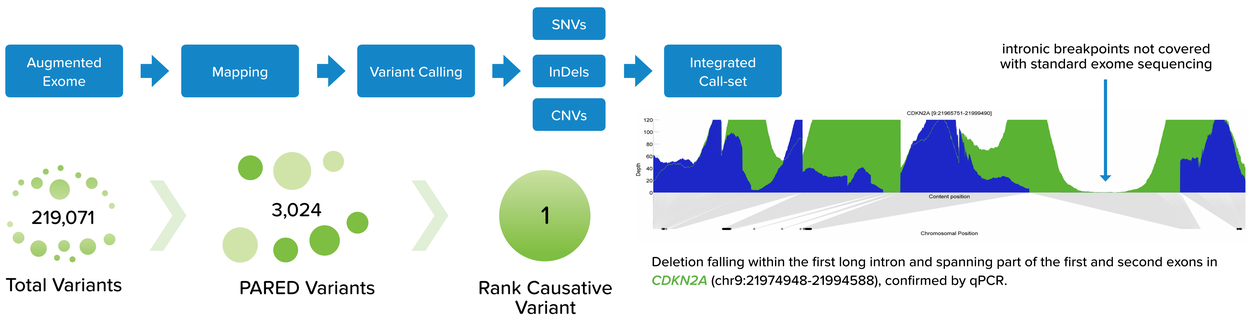


50% reciprocal overlap (>1kB)	sensitivity, genome-wide	sensitivity, medically relevant	FP/TP
40x WGS (133G)	99%	88%	27
Augmented Exome (12G)	63%	82%	5

Among medically reported variants in the gold set sensitivity with the ACE Augmented Exome is on-par with 40x WGS despite a >10x reduction in sequencing data. ACE shows ~5-fold reduction in the number of FPs for every TP detected compared to WGS.

Clinical Case

Proband had Extensive Family History of Neurofibromatosis with Melanoma, Neural Tumors



Deletion falling within the first long intron and spanning part of the first and second exons in *CDKN2A* (chr9:21974948-21994588), confirmed by qPCR.

[2]Kim *et al.*, Wessim: a whole-exome sequencing simulator based on in silico exome capture. Bioinformatics. 29(8): 1076-7–251 (2014)

www.personalis.com | info@personalis.com | +1 855-436-6634 | +1 650-752-1300

POS-AMAP-1.0 NOV 2014

Personalis®
Pioneering Genome-Guided Medicine