

Creation of an Expert Curated Variant List for Clinical Genomic Test Development and Validation: A ClinGen and GeT-RM Collaborative Project

Short Running Title (40 characters): Creation of a Curated Variant List

Emma Wilcox^{1*}, Steven M. Harrison^{1*}, Edward Lockhart², Karl Voelkerding³, Ira M. Lubin⁴, ClinGen Expert Panels⁵ Heidi L. Rehm^{1,6^}, Lisa Kalman^{7^}, Birgit Funke^{8^}

*, ^ these authors contributed equally

¹Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA
(ewilcox@broadinstitute.org, sharriso@broadinstitute.org, hrehm@broadinstitute.org)

²Informatics and Data Science Branch, Division of Laboratory Systems, Centers for Disease Control and Prevention, Atlanta, GA coy3@cdc.gov

³Voelkerding Consulting, Salt Lake City, UT

⁴Quality and Safety Systems Branch, Division of Laboratory Systems, Centers for Disease Control and Prevention, Atlanta, GA iblo@cdc.gov

⁵See Supplemental Table 1

⁶Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA

⁷Informatics and Data Science Branch, Division of Laboratory Systems, Centers for Disease Control and Prevention, Atlanta, GA LJK0@cdc.gov

⁸Division of Genomic Health, Sema4, Stamford, CT

Corresponding Author:

Lisa Kalman, PhD
Informatics and Data Science Branch
Division of Laboratory Systems
Office of Surveillance, Epidemiology, and Laboratory Services
Centers for Disease Control and Prevention
1600 Clifton Road
Mailstop V24-3
Atlanta, GA 30333
(404) 498-2231
LKalman@cdc.gov

Disclosures:

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention/the Agency for Toxic Substances and Disease Registry or the National Institutes of Health. Use of trade names and commercial sources is for identification only and does not imply endorsement by the Centers for Disease Control and Prevention, the Public Health Service, the US Department of Health and Human Services or the National Institutes of Health.

Acknowledgements:

40 ClinGen is primarily funded by the National Human Genome Research Institute (NHGRI), through the following
41 three grants: U41HG006834, U41HG009649, U41HG009650. ClinGen also receives support for content curation
42 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD), through
43 the following three grants: U24HD093483, U24HD093486, U24HD093487.
44

45 **Abstract (220 words max)**

46 Modern genomic sequencing tests often interrogate large numbers of genes. Identification of appropriate
47 reference materials for development, validation studies, and quality assurance of these tests poses a significant
48 challenge for laboratories. It is difficult to develop and maintain expert knowledge to identify all variants that
49 must be validated to assure analytic and clinical validity. Additionally, it is usually not possible to procure
50 appropriate and characterized genomic DNA reference materials containing the number and scope of variants
51 required. To address these challenges, the Centers for Disease Control and Prevention's Genetic Testing
52 Reference Material Program (GeT-RM) has partnered with the Clinical Genome Resource (ClinGen) to develop a
53 publicly available list of expert curated, clinically important variants. ClinGen Variant Curation Expert Panels
54 nominated 546 variants found in 84 disease associated genes, including common pathogenic and difficult to
55 detect variants. Variant types nominated included 346 SNVs, 104 deletions, 37 CNVs, 25 duplications, 18
56 deletion-insertions, 5 inversions, 4 insertions, 2 complex rearrangements, 3 in difficult to sequence regions, and
57 2 fusions. This expert-curated variant list is a resource that provides a foundation for designing comprehensive
58 validation studies and for creating *in silico* reference materials for clinical genomic test development and
59 validation.
60

61 **Introduction**

62 Genetic testing has evolved from interrogating small sets of known, pathogenic variants in one or a few genes
63 using targeted genotyping assays or Sanger sequencing to examining hundreds or thousands of genes at a time
64 using next generation sequencing (NGS). Large gene panels are now the norm in many clinical areas, such as
65 hereditary cancer and cardiomyopathy, and are particularly useful for disorders with locus and allelic

66 heterogeneity. Offerings from molecular testing laboratories range from single disorder
67 panels, the latter often allowing molecular refinement of the initial clinical diagnosis (National Institutes of
68 Health, Genetic Testing Registry <https://www.ncbi.nlm.nih.gov/gtr/> accessed 5/6/2021).

69
70 While NGS has advanced many aspects of genetic testing, this technology has made the task of designing and
71 developing analytically and clinically valid tests more complex. This process requires deep clinical and genetic
72 expertise to understand the clinically relevant variant spectrum, and characterization of genomic context as high
73 GC content, highly homologous genes, or repetitive sequences pose challenges for most current NGS
74 technologies.^{1, 2} When examining only one or a few well-studied genes, sufficient expertise to identify clinically
75 important variants can be developed easily, but this aspect of test design and development does not scale to
76 assays that examine hundreds or thousands of genes.³ Laboratories often cannot procure reference materials
77 (RMs) that encompass the scope of variants and variant types needed for NGS test development and validation
78 because the supply of available characterized genomic DNA samples from cell lines or patient samples is limited
79 and does not cover the range of genes and variants often included in NGS assays. Finally, traditional patient
80 derived RMs, such as genomic DNA from cell lines, typically contain only 1 or 2 clinically important variants per
81 sample, which can significantly increase the number of samples needed, and thus the complexity and cost of a
82 validation study when large numbers of variants need to be addressed.

83
84 Proper analytical validation of genomic tests and large test panels requires a combination of approaches: To
85 ensure optimal confidence in the ability of a given technology to detect variants within a genome or subregion
86 (e.g., exon), large numbers of variants for each clinically relevant type (SNVs, insertions, deletions, SVs) need to
87 be analyzed. In addition, test developers should demonstrate the test's ability to detect specific clinically
88 important variants particularly when they are technically challenging (Food and Drug Administration
89 <https://www.fda.gov/media/99208/download>, accessed 3/23/2021; College of American Pathologists,

90 Molecular Pathology 2020 checklist <https://www.cap.org/> registration required, accessed 3/23/2021).⁴

91 Different types of reference materials have been created to meet these needs.

92

93 The 1000 Genomes project HapMap samples (1000 Genomes Project

94 <https://www.internationalgenome.org/home>; accessed 10/11/2020) were the first generation of characterized

95 genomes that were publicly available. More recently, the National Institute for Standards and Technology (NIST)

96 Genome in a Bottle project has characterized 7 publicly available genomes including NA12878 and two

97 son/mother/father trios of Ashkenazi Jewish and Han Chinese ancestry for use as reference materials (NIST

98 <https://www.nist.gov/programs-projects/genome-bottle> accessed 5/10/2021).^{5, 6} These samples have been

99 extensively characterized using a number of different technologies and variant calling methods covering over

100 90% of the genome. NIST has generated publicly available, high confidence benchmarking data sets that have

101 been widely used by the clinical testing community and are now also covering structural variants, and variants in

102 highly homologous regions.⁷ In addition, Illumina produced a set of over 4.7 million phased variants from 17

103 samples of a large multigenerational pedigree.⁸ These data sets are excellent resources for establishing general

104 analytical performance across various variant types, but do not meet the additional need to validate specific

105 clinically important variants. Other publicly available genomic DNA samples, such as those characterized by the

106 Center for Disease Control and Prevention's Genetic Testing Reference Material Program (GeT-RM

107 <https://www.cdc.gov/labquality/get-rm/index.html> accessed 10/11/2020) and samples available from

108 repositories such as Coriell Institute for Medical Research (Coriell <https://www.coriell.org/> accessed 5/10/2021)

109 or the American Type Culture Collection (ATCC) (ATCC <https://www.atcc.org/> accessed 5/10/2021) have

110 identified variants in some, but not all genes likely to be included in NGS tests. Although useful, none of these

111 genomic DNA samples, individually or as a group, are sufficient to represent the wide range of variants needed

112 for a thorough and comprehensive validation of an NGS assay.

114 As NGS tests have grown to include hundreds to thousands of genes, it has become technically and economically
115 impractical to rely exclusively on DNA samples for assay development and validation studies to assess assay
116 performance. Procuring sufficient reference materials containing specific variants has always been challenging
117 and can be virtually impossible for large gene panels, particularly those that span multiple disorders. In addition,
118 the commoditization of genomic testing and the availability of “off-the-shelf” gene panels have facilitated
119 implementation of genomic tests. -Well established laboratories with longstanding clinical expertise often have
120 archived DNA which can be used as reference materials for test development. In contrast, laboratories that do
121 not already have in-depth expertise in the genomic profile of the disorders included in the panels are faced with
122 a paucity of organized resources and guidance as to which clinically important variants require specific attention
123 during test development. The FDA and professional organizations including the College of American Pathologists
124 have begun to acknowledge these challenges and have responded with guidance that permits use of *in silico*
125 reference materials to supplement clinical test validation and operation (Food and Drug Administration
126 <https://www.fda.gov/media/99208/download>, accessed 3/23/2021; College of American Pathologists,
127 Molecular Pathology 2020 checklist <https://www.cap.org/> registration required, accessed 3/23/2021).

128

129 In 2019, the Clinical Laboratory Improvement Advisory Committee (CLIAAC) issued a recommendation to the
130 Department of Health and Human Services that the GeT-RM program should develop electronic reference
131 materials for NGS testing (https://www.cdc.gov/cliac/docs/summary/cliac0419_summary.pdf accessed
132 [5/10/2021](https://www.cdc.gov/cliac/docs/summary/cliac0419_summary.pdf)). As a first step towards this goal, GeT-RM has partnered with the Clinical Genome Resource
133 (ClinGen, <https://www.clinicalgenome.org/> accessed 2/27/2021) to develop lists of expert curated, clinically
134 important variants. This publication describes these variant lists, which can be used to inform test validation
135 studies and will serve as a foundation for creating multi-variant electronic reference materials by *in silico*
136 mutagenesis of laboratory generated BAM or FASTQ files.⁹⁻¹¹

137

138

139 **Methods**

140 Variant nomination process

141 Members of 36 expert panels, including 35 of ClinGen's Variant Curation Expert Panels (VCEPs), and experts
142 from the CFTR2 Project were contacted in March 2020 and asked to nominate clinically important variants in the
143 genes covered by the expert panels. FDA has recognized ClinGen's curation process and its resulting variant
144 classifications as a regulatory-grade variant database (ClinGen <https://clinicalgenome.org/docs/fda-recognizes-clingen-assertions-in-clinvar-frequently-asked-questions/> accessed 5/10/2021; Food and Drug Administration
145 <https://www.fda.gov/news-events/press-announcements/fda-takes-new-action-advance-development-reliable-and-beneficial-genetic-tests-can-improve-patient> accessed 5/10/2021). These VCEP curated, FDA recognized
146 variants are available via NCBI's ClinVar database and ClinGen's Evidence Repository
147 (<https://erepo.clinicalgenome.org> accessed 5/10/2021).

150

151 Each expert panel was asked to nominate a minimum of 10 variants per disease area, including pathogenic
152 variants that are the largest contributors to a disease as well as variants that may represent potential analytic
153 challenges to laboratories. Panels submitted their variant nominations on a standardized template from April-
154 June 2020. For each variant, the panels noted if the basis for inclusion was: a) Major contributor to disease, b)
155 Analytic detection difficult, c) Filtration impact (may be inadvertently filtered out due to high allele frequency),
156 or d) Other (with explanation provided in a comment box). The panels were instructed that any reason for
157 difficult detection was appropriate including: variant type, homology issues or a high allele frequency for a lower
158 penetrant variant that might make the variant excluded by standard allele frequency filters. The ClinGen Allele
159 Registry was used to standardize nomenclature for all nominated variants and ClinVar Variation IDs and
160 associated disorders were added where available. A full list of the expert panels that contributed variants can be
161 found in Supplemental Table 1 and the full list of nominated variants can be found in Supplemental Table 2.

162

163 Not all of the ClinGen VCEPs that participated in this project have completed all four steps of the VCEP
164 development process required for final approval by ClinGen¹² (ClinGen
165 <https://clinicalgenome.org/docs/guidelines-for-applying-for-variant-or-gene-curation-expert-panel-status/>
166 accessed 3/23/2021). Of the variants submitted by the panels, 38% are from expert panels with final ClinGen
167 approval to submit to ClinVar as ‘reviewed by expert panel’, whereas 62% are from ClinGen VCEPs that were still
168 developing disease specifications for variant classification and were not yet fully ClinGen-approved at the time
169 of submission of variants to the project. VCEP status at the time of manuscript submission is shown in column B
170 of Supplemental table 2. Current status can be found on the ClinGen website
171 (<https://clinicalgenome.org/working-groups/clinical-domain/> accessed 3/23/2021).

172

173 Limitations of variant selection process

174 Several limitations of the variant selection process should be noted. There were no explicit requirements given
175 to the expert panels in terms of the definitions of categories a, b, and c and therefore each may have used
176 slightly different definitions to select variants to nominate for this project. In addition, some expert panels may
177 have used laboratory data that is not in the public domain to identify nominated variants and this information
178 was not requested during the submission process. Furthermore, some expert panels represent genes and
179 diseases that are very rare and therefore variants may not, in general, be highly prevalent. These are not
180 considered serious limitations as these are “first-generation” variant lists, which provide an entirely new
181 resource. It is expected that this effort will continue in years to come to broaden and deepen this resource.

182

183 **Results**

184 Nominations of clinically important pathogenic or difficult to detect variants were received from 24 expert
185 panels out of 36 contacted (Supplemental Table 1). Overall, 546 unique variants in 84 genes were nominated
186 (Supplemental Table 2) including 22 of the 59 genes recommended by ACMG for reporting of incidental or

187 secondary findings, which are included as part of virtually all proactive genomic health tests.¹² The nominated
188 variants are causative for a wide range of diseases many of which are commonly tested by NGS including
189 heritable cancers, inborn errors of metabolism, cardiomyopathy, diabetes, and immune disorders (Table 1). A
190 number of different variant types were nominated, including 346 SNVs, 104 deletions, 37 CNVs, 25 duplications,
191 18 deletion-insertions, 5 inversions, 4 insertions, 2 complex rearrangements, 2 fusions, and 3 variants in regions
192 defined as “difficult region to cover and map variants”. The variants were nominated for a variety of reasons:
193 355 are major contributors to disease, 111 are analytically difficult to detect, 28 may have filtration impacts (be
194 inadvertently filtered out due to high allele frequency), 6 had multiple reasons for inclusion, and 46 were listed
195 as “other” or no explanation was provided (Supplemental Table 2).

196

197 Discussion

198 Clinical laboratories in the United States are required by regulation and guided by professional or best practice
199 standards to use characterized reference materials for test development, validation and verification studies,
200 quality control and proficiency testing¹³⁻¹⁶ (American College of Medical Genetics
201 <https://www.acmg.net/PDFLibrary/Standards-Guidelines-Clinical-Molecular-Genetics.pdf> , accessed 4/9/2020,
202 Washington State Legislature, <http://app.leg.wa.gov/WAC/default.aspx?cite=246-338-090> , accessed 4/9/2020,
203 College of American Pathologists <https://www.cap.org/> , accessed 4/9/2020 (registration required), New York
204 State Clinical Laboratory Evaluation Program, <http://www.wadsworth.org/clep>, accessed 4/9/2020).

205

206 Certain aspects of test development and validation, such as DNA extraction, library preparation, and sequence
207 generation will always require the use of genomic DNA from cell lines or patient samples. However, given the
208 scope of many NGS tests, it is no longer feasible to rely solely upon genomic DNA samples containing only one or
209 a few variants of interest to validate an assay that tests an entire exome or genome or many genes within a
210 large NGS panel, and novel types of reference materials are needed to enable assessment of a test’s analytical
211 validity and establish adequate clinical validity.

212

213 The Association for Molecular Pathology and the College of American Pathologists allow the use of *in silico*
214 reference materials to supplement genomic DNA for the validation of bioinformatic pipelines (College of
215 American Pathologists, Molecular Pathology 2020 checklist <https://www.cap.org/> registration required,
216 accessed 3/23/2021).^{4, 17-19} Specialized software^{9, 10, 20} has been developed to edit sequence variants into BAM
217 or FASTQ files generated by the assay being validated to create customized reference materials. The software is
218 publicly available, and several companies offer this service to laboratories. The GeT-RM program is planning to
219 use these curated variants and file editing software to pilot development and use of *in silico* reference materials
220 in clinical laboratories.

221

222 The number of variants that can be added is potentially unlimited, but care must be taken to avoid adding too
223 many variants in *cis* to the same sequence reads because it may affect alignment to the reference sequence. The
224 types of variants that can be introduced remains limited to SNVs and smaller insertions and deletions at
225 different allele fractions, although the ability to introduce CNVs is currently being developed. This approach has
226 been used successfully by the College of American Pathologists for its NGS proficiency testing challenges.^{9, 10}

227

228 The variant list developed by the effort presented here is available on the GeT-RM and ClinGen websites and can
229 serve as a knowledge resource to help laboratories identify important pathogenic and difficult to detect variants
230 in genes that are included in their assays, as well as a foundation for generating customized *in silico* reference
231 materials. GeT-RM and ClinGen will continue to collaborate to add to the current variant list and invite input
232 from the genetics community about this list and the processes used to generate it.

233

234

235

236

| Gene | HGNC Gene ID | # Variants nominated | Disorder |
|--------|--------------|-------------------------|---------------------------------------------------------------------------------|
| ACADVL | HGNC:92 | 10 | Very long-chain acyl-CoA dehydrogenase deficiency |
| ACVRL1 | HGNC:175 | 3 | Hereditary hemorrhagic telangiectasia, type 2 |
| APC | HGNC:583 | 4 | Familial adenomatous polyposis |
| ATM | HGNC:795 | 6 | Hereditary breast carcinoma (AD) Ataxia telangiectasia (AR) |
| BARD1 | HGNC:952 | 5 | Hereditary breast carcinoma |
| BRAF | HGNC:1097 | 1 | Cardiofaciocutaneous syndrome |
| BRCA1 | HGNC:1100 | 26 | Breast-ovarian cancer, familial, susceptibility to, 1 (AD) |
| BRCA2 | HGNC:1101 | 22 | Breast-ovarian cancer, familial, susceptibility to, 2 (AD) |
| BRIP1 | HGNC:20473 | 5 | Familial ovarian cancer (AD) Fanconi anemia complementation group J (AR) |
| BTK | HGNC:1133 | 10 | X-linked agammaglobulinemia (XLA) |
| CARD11 | HGNC:16393 | 5 | CARD11 defects |
| CDH1 | HGNC:1748 | 10 | Hereditary diffuse gastric adenocarcinoma |

| | | | |
|---------|------------|----|---------------------------------------------------------------------------------|
| CFH | HGNC:4883 | 5 | Atypical hemolytic uremic syndrome (aHUS)/Thrombotic microangiopathy (TMA) |
| CFTR | HGNC:1884 | 14 | Cystic fibrosis |
| CHEK2 | HGNC:16627 | 5 | Hereditary breast carcinoma |
| COL6A1 | HGNC:2211 | 1 | Congenital muscular dystrophy |
| CYP1B1 | HGNC:2597 | 3 | Congenital glaucoma |
| DCLRE1C | HGNC:17642 | 8 | Severe combined immunodeficiency due to DCLRE1C deficiency |
| DES | HGNC:2770 | 17 | Arrhythmogenic right ventricular cardiomyopathy (arrhythmogenic cardiomyopathy) |
| DICER1 | HGNC:17098 | 7 | Pleuropulmonary blastoma |
| DNM2 | HGNC:2974 | 3 | Centronuclear myopathy |
| DSC2 | HGNC:3036 | 5 | Arrhythmogenic right ventricular cardiomyopathy (arrhythmogenic cardiomyopathy) |
| DSG2 | HGNC:3049 | 7 | Arrhythmogenic right ventricular cardiomyopathy (arrhythmogenic cardiomyopathy) |
| DSP | HGNC:3052 | 10 | Arrhythmogenic right ventricular cardiomyopathy (arrhythmogenic cardiomyopathy) |

| | | | |
|-------|-----------|----|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| ENG | HGNC:3349 | 4 | Hereditary hemorrhagic telangiectasia |
| FBN1 | HGNC:3603 | 10 | Marfan syndrome, familial thoracic aortic aneurysm and aortic dissection |
| GAA | HGNC:4065 | 8 | Glycogen storage disease type II |
| GATA2 | HGNC:4171 | 15 | GATA2 deficiency with susceptibility to MDS/AML; GATA2 haploinsufficiency |
| GCK | HGNC:4195 | 11 | Monogenic diabetes |
| HBA1* | HGNC:4823 | 4 | Alpha thalassemia, Hemoglobin H disease, Hemoglobin Bart's hydrops fetalis |
| HBA2* | HGNC:4824 | 7 | Alpha thalassemia, Hemoglobin H disease, Hemoglobin Bart's hydrops fetalis |
| HBB* | HGNC:4827 | 10 | Sickle cell anemia, sickle cell disease associated with another hemoglobin anomaly, Beta thalassemia, Sickle cell-beta-thalassemia disease syndrome, Sickle cell-hemoglobin C disease syndrome, Hemoglobin C-beta-thalassemia syndrome, Hemoglobin C disease, Sickle cell-hemoglobin D disease syndrome, Hemoglobin D disease, Sickle cell-hemoglobin E disease syndrome, Hemoglobin E-beta-thalassemia syndrome, Hemoglobin E disease, Hemoglobin Lepore-beta-thalassemia syndrome |

| | | | |
|--------|------------|----|--------------------------------------------------------------------------------------------|
| HBD* | HGNC:4829 | 1 | Hemoglobin Lepore-beta-thalassemia syndrome, sickle cell-beta-thalassemia disease syndrome |
| HBQ1 | HGNC:4833 | 1 | Alpha thalassemia, Hemoglobin H disease, Hemoglobin Bart's hydrops fetalis |
| HNF1A | HGNC:11621 | 2 | Monogenic diabetes |
| HNF1B | HGNC:11630 | 1 | Monogenic diabetes |
| HNF4A | HGNC:5025 | 3 | Monogenic diabetes |
| HRAS | HGNC:5173 | 1 | Costello syndrome |
| IKBKG | HGNC:5961 | 6 | NEMO (NFkB essential modulator) deficiency |
| IL2RG | HGNC:6010 | 10 | SCID (severe combined immunodeficiency) |
| INS | HGNC:6081 | 1 | Monogenic diabetes |
| JAK3 | HGNC:6193 | 5 | SCID (severe combined immunodeficiency) |
| LDLR | HGNC:6547 | 10 | Familial hypercholesterolemia |
| LRBA | HGNC:1742 | 5 | LRBA deficiency |
| LYST | HGNC:1968 | 5 | Chediak-Higashi syndrome (CHS) |
| MAP2K1 | HGNC:6840 | 1 | Cardiofaciocutaneous syndrome |
| MLH1 | HGNC:7127 | 19 | Colorectal cancer, hereditary nonpolyposis, type 2 |

| | | | |
|--------|------------|----|---------------------------------------------------------------------------------|
| | | | Constitutional mismatch repair deficiency syndrome |
| MSH2 | HGNC:7325 | 16 | Lynch syndrome |
| MSH6 | HGNC:7329 | 7 | Lynch syndrome |
| MUTYH | HGNC:7527 | 2 | MUTYH-related attenuated familial adenomatous polyposis (AR/AD) |
| MYBPC3 | HGNC:7551 | 16 | Hypertrophic cardiomyopathy |
| MYH7 | HGNC:7577 | 6 | Hypertrophic cardiomyopathy |
| MYOC | HGNC:7610 | 4 | Open angle glaucoma |
| NCF1 | HGNC:7660 | 6 | Chronic Granulomatous Disease (CGD) |
| NEB | HGNC:7720 | 2 | Nemaline myopathy |
| NTHL1 | HGNC:8028 | 1 | Familial adenomatous polyposis |
| PAH | HGNC:8582 | 17 | Phenylketonuria |
| PALB2 | HGNC:26144 | 5 | Hereditary breast carcinoma (AD), Fanconi anemia complementation group N (AR) |
| PKP2 | HGNC:9024 | 22 | Arrhythmogenic right ventricular cardiomyopathy (arrhythmogenic cardiomyopathy) |
| PLN | HGNC:9080 | 4 | Arrhythmogenic right ventricular cardiomyopathy (arrhythmogenic cardiomyopathy) |
| POLD1 | HGNC:9175 | 1 | Polymerase proofreading-related adenomatous polyposis |
| POLE | HGNC:9177 | 1 | Polymerase proofreading-related adenomatous polyposis |

| | | | |
|---------|------------|----|------------------------------------------------------------------------------------|
| PTEN | HGNC:9588 | 5 | PTEN hamartoma tumor syndrome |
| PTPN11 | HGNC:9644 | 4 | Noonan syndrome, Noonan syndrome with multiple lentigines |
| RAD51C | HGNC:9820 | 5 | Familial ovarian cancer (AD) |
| RAD51D | HGNC:9823 | 5 | Familial ovarian cancer (AD) |
| RAF1 | HGNC:9829 | 1 | Noonan syndrome |
| RAG1 | HGNC:9831 | 7 | SCID (severe combined immunodeficiency) |
| RUNX1 | HGNC:10471 | 10 | Familial platelet disorder with associated myeloid malignancy |
| RYR1** | HGNC:10483 | 17 | Susceptibility to malignant hyperthermia (AD) Central core myopathy (AR/AD) |
| SELENON | HGNC:15999 | 4 | Rigid spine myopathy |
| SHOC2 | HGNC:15454 | 1 | Noonan syndrome-like disorder with loose anagen hair |
| SMAD4 | HGNC:6770 | 3 | Juvenile polyposis/hereditary hemorrhagic telangiectasia syndrome |
| SOS1 | HGNC:11187 | 1 | Noonan syndrome |
| TMEM43 | HGNC:28472 | 1 | Arrhythmogenic right ventricular cardiomyopathy (arrhythmogenic cardiomyopathy) |

| | | | |
|---------------------------------------------------------------------------------------------------------------------------|------------|----|---------------------------------------------------------------------------------------------------------|
| TNNI3 | HGNC:11947 | 1 | Hypertrophic cardiomyopathy |
| TNNT2 | HGNC:11949 | 2 | Hypertrophic cardiomyopathy, dilated cardiomyopathy |
| TP53 | HGNC:11998 | 10 | Li-Fraumeni syndrome |
| TPM2 | HGNC:12011 | 1 | Nemaline myopathy |
| TPM3 | HGNC:12012 | 3 | Nemaline myopathy |
| TTN | HGNC:12403 | 7 | Distal titinopathy, tibial muscular dystrophy, hereditary myopathy with early respiratory failure |
| VHL | HGNC:12687 | 7 | Von Hippel-Lindau disease |
| WAS | HGNC:12731 | 13 | Wiskott- Aldrich syndrome (WAS) |
| WFS1 | HGNC:12762 | 1 | Monogenic diabetes |
| * Some variants in these genes are multi-gene deletions or fusions. | | | |
| **One variant in RYR1 was submitted by both the Congenital Myopathies and Malignant Hyperthermia Susceptibility VCEPs. | | | |

238

239 Suppl Table 1. List of participating expert panels and names of participants

240 Suppl Table 2. List of variants nominated

241

242 References

- 243 1. Mandelker D, Schmidt RJ, Ankala A, McDonald Gibson K, Bowser M, Sharma H, Duffy E, Hegde M,
244 Santani A, Lebo M, Funke B: Navigating highly homologous genes in a molecular diagnostic setting: a
245 resource for clinical next-generation sequencing. Genetics in medicine : official journal of the American
246 College of Medical Genetics 2016, 18:1282-1289.

2. Bean LJH, Funke B, Carlston CM, Gannon JL, Kantarci S, Krock BL, Zhang S, Bayrak-Toydemir P, Committee ALQA: Diagnostic gene sequencing panels: from design to report-a technical standard of the American College of Medical Genetics and Genomics (ACMG). *Genetics in medicine : official journal of the American College of Medical Genetics* 2020, 22:453-461.
3. Rehm HL, Berg JS, Brooks LD, Bustamante CD, Evans JP, Landrum MJ, Ledbetter DH, Maglott DR, Martin CL, Nussbaum RL, Plon SE, Ramos EM, Sherry ST, Watson MS, ClinGen: ClinGen--the Clinical Genome Resource. *The New England journal of medicine* 2015, 372:2235-2242.
4. Clinical and Laboratory Standards Institute: MM17 Validation and Verification of Multiplex Nucleic Acid Assays, 2nd ed. Wayne, PA: Clinical and Laboratory Standards Institute, 2018.
5. Zook JM, Catoe D, McDaniel J, Vang L, Spies N, Sidow A, Weng Z, Liu Y, Mason CE, Alexander N, Henaff E, McIntyre ABR, Chandramohan D, Chen F, Jaeger E, Moshrefi A, Pham K, Stedman W, Liang T, Saghibini M, Dzakula Z, Hastie A, Cao H, Deikus G, Schadt E, Sebra R, Bashir A, Truty RM, Chang CC, Gulbahce N, Zhao K, Ghosh S, Hyland F, Fu Y, Chaisson M, Xiao C, Trow J, Sherry ST, Zaranek AW, Ball M, Bobe J, Estep P, Church GM, Marks P, Kyriazopoulou-Panagiotopoulou S, Zheng GXY, Schnall-Levin M, Ordonez HS, Mudivarti PA, Giorda K, Sheng Y, Rypdal KB, Salit M: Extensive sequencing of seven human genomes to characterize benchmark reference materials. *Scientific Data* 2016, 3:160025.
6. Wang Y-C, Olson ND, Deikus G, Shah H, Wenger AM, Trow J, Xiao C, Sherry S, Salit ML, Zook JM, Smith M, Sebra R: High-coverage, long-read sequencing of Han Chinese trio reference samples. *Scientific Data* 2019, 6:91.
7. Zook JM, McDaniel J, Olson ND, Wagner J, Parikh H, Heaton H, Irvine SA, Trigg L, Truty R, McLean CY, De La Vega FM, Xiao C, Sherry S, Salit M: An open resource for accurately benchmarking small variant and reference calls. *Nature biotechnology* 2019, 37:561-566.
8. Eberle MA, Fritzilas E, Krusche P, Kallberg M, Moore BL, Bekritsky MA, Iqbal Z, Chuang HY, Humphray SJ, Halpern AL, Kruglyak S, Margulies EH, McVean G, Bentley DR: A reference data set of 5.4 million phased human variants validated by genetic inheritance from sequencing a three-generation 17-member pedigree. *Genome Res* 2017, 27:157-164.
9. Duncavage EJ, Abel HJ, Pfeifer JD: In Silico Proficiency Testing for Clinical Next-Generation Sequencing. *The Journal of molecular diagnostics : JMD* 2017, 19:35-42.
10. Duncavage EJ, Abel HJ, Merker JD, Bodner JB, Zhao Q, Voelkerding KV, Pfeifer JD: A Model Study of In Silico Proficiency Testing for Clinical Next-Generation Sequencing. *Archives of pathology & laboratory medicine* 2016, 140:1085-1091.
11. Li Z, Fang S, Zhang R, Yu L, Zhang J, Bu D, Sun L, Zhao Y, Li J: VarBen: Generating in Silico Reference Data Sets for Clinical Next-Generation Sequencing Bioinformatics Pipeline Evaluation. *The Journal of molecular diagnostics : JMD* 2021, 23:285-299.
12. Kalia SS, Adelman K, Bale SJ, Chung WK, Eng C, Evans JP, Herman GE, Hufnagel SB, Klein TE, Korf BR, McKelvey KD, Ormond KE, Richards CS, Vlangos CN, Watson M, Martin CL, Miller DT, on behalf of the ASFMWG: Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update (ACMG SF v2.0): a policy statement of the American College of Medical Genetics and Genomics. *Genetics in Medicine* 2017, 19:249-255.
13. Chen B, Gagnon M, Shahangian S, Anderson NL, Howerton DA, Boone JD, Centers for Disease C, Prevention: Good laboratory practices for molecular genetic testing for heritable diseases and conditions. *MMWR Recommendations and reports : Morbidity and mortality weekly report Recommendations and reports / Centers for Disease Control* 2009, 58:1-37; quiz CE-31-34.
14. Association for Molecular Pathology statement. Recommendations for in-house development and operation of molecular diagnostic tests. *Am J Clin Pathol* 1999, 111:449-463.
15. International Organization for Standardization: ISO 15189 Medical Laboratories: Requirements for Quality and Competence. Geneva. International Organization for Standardization 2012.
16. Electronic Code of Federal Regulations: The Clinical Laboratory Improvement Amendments (CLIA). Title 42 Chapter IV, Subchapter G, Part 493.

296 17. Aziz N, Zhao Q, Bry L, Driscoll DK, Funke B, Gibson JS, Grody WW, Hegde MR, Hoeltge GA, Leonard DG,
297 Merker JD, Nagarajan R, Palicki LA, Robetorye RS, Schrijver I, Weck KE, Voelkerding KV: College of
298 American Pathologists' laboratory standards for next-generation sequencing clinical tests. Archives of
299 pathology & laboratory medicine 2015, 139:481-493.

300 18. Roy S, Coldren C, Karunamurthy A, Kip NS, Klee EW, Lincoln SE, Leon A, Pullambhatla M, Temple-Smolkin
301 RL, Voelkerding KV, Wang C, Carter AB: Standards and Guidelines for Validating Next-Generation
302 Sequencing Bioinformatics Pipelines: A Joint Recommendation of the Association for Molecular
303 Pathology and the College of American Pathologists. The Journal of molecular diagnostics : JMD 2018,
304 20:4-27.

305 19. Santani A, Simen BB, Briggs M, Lebo M, Merker JD, Nikiforova M, Vasalos P, Voelkerding K, Pfeifer J,
306 Funke B: Designing and Implementing NGS Tests for Inherited Disorders: A Practical Framework with
307 Step-by-Step Guidance for Clinical Laboratories. The Journal of molecular diagnostics : JMD 2019,
308 21:369-374.

309 20. Patil SA, Mujacic I, Ritterhouse LL, Segal JP, Kadri S: insiM: in silico Mutator Software for Bioinformatics
310 Pipeline Validation of Clinical Next-Generation Sequencing Assays. The Journal of molecular diagnostics :
311 JMD 2019, 21:19-26.