

Systematic large-scale application of ClinGen InSiGHT APC-specific ACMG/AMP variant classification criteria substantially alleviates the burden of variants of uncertain significance in ClinVar and LOVD databases

Xiaoyu Yin,^{1,2,3} Marcy Richardson,⁴ Andreas Laner,⁵ Xuemei Shi,⁶ Elisabet Ognedal,⁷ Valeria Vasta,⁸ Thomas v. O. Hansen,^{9,10} Marta Pineda,^{11,12,13} Deborah Ritter,^{14, 15} Johan T. den Dunnen,¹⁶ Emadeldin Hassanin,^{17,18} Wencong Lyman Lin,¹⁹ Ester Borrás,²⁰ Karl Krahn,²¹ Margareta Nordling,^{22,23} Alexandra Martins,²⁴ Khalid Mahmood,²⁵ Emily A.W. Nadeau,²⁶ Victoria Beshay,²⁷ Carli Tops,¹⁶ Maurizio Genuardi,²⁸ Tina Pesaran,⁴ Ian M. Frayling,^{29,30,31} Gabriel Capellá,^{11,12,13} Andrew Latchford,^{29,32} Sean V. Tavtigian,^{33,34} Carlo Maj,^{17,35} Sharon E. Plon,^{14,15} Marc S. Greenblatt,²⁶ Finlay A. Macrae,^{1,2} Isabel Spier,^{3,11,36*} Stefan Aretz^{3,11,36*}

*These authors contributed equally to this study

¹ Department of Colorectal Medicine and Genetics, Royal Melbourne Hospital, Parkville, Australia

² Department of Medicine, University of Melbourne, Parkville, Australia

³ Institute of Human Genetics, Medical Faculty, University of Bonn, Bonn, Germany

⁴ Ambry Genetics, Aliso Viejo, California, USA

⁵ Medical Genetics Center Munich, MGZ Munich, Germany

⁶ Greenwood Genetic Center, Greenwood, South Carolina, USA

⁷ Western Norway Familial Cancer Center, Haukeland University Hospital, Bergen, Norway

⁸ Northwest Genomics Center, Department of Genome Sciences, University of Washington, Seattle

⁹ Department of Clinical Genetics, Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark

¹⁰ Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark

¹¹ European Reference Network on Genetic Tumour Risk Syndromes (ERN GENTURIS) – Project ID No 739547

¹² Hereditary Cancer Program, Catalan Institute of Oncology – ONCOBELL, IDIBELL, Barcelona, Spain

¹³ Centro de Investigación Biomédica en Red de Cáncer (CIBERONC), Instituto Salud Carlos III, Madrid, Spain

¹⁴ Baylor College of Medicine, Houston, Texas, USA

¹⁵ Texas Children's Cancer Center, Texas Children's Hospital, Houston, Texas, USA

¹⁶ Departments of Human Genetics & Clinical Genetics, Leiden University Medical Center, Leiden, Netherlands

¹⁷ Institute for Genomic Statistics and Bioinformatics, University Hospital Bonn, Bonn, Germany

¹⁸ Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg

¹⁹ St. Vincent's Hospital Melbourne, East Melbourne, Australia

²⁰ Invitae Corporation, San Francisco, California, USA

²¹ GeneDx, Gaithersburg, Maryland, USA

²² Department of Biomedical and Clinical Sciences, Linköping University, Linköping, Sweden

²³ Department of Clinical Genetics, Linköping University Hospital, Linköping, Sweden

²⁴ Université de Rouen Normandie, Inserm U1245, F-76000 Rouen, France

²⁵ Colorectal Oncogenomics Group, Department of Clinical Pathology, University of Melbourne, Melbourne, Australia

²⁶ Department of Medicine, Larner College of Medicine, University of Vermont, Burlington, Vermont, USA

²⁷ Peter MacCallum Cancer Centre, Melbourne, Australia

²⁸ Fondazione Policlinico Universitario A. Gemelli IRCCS, and Dipartimento di Scienze della Vita e Sanità Pubblica, Università Cattolica del Sacro Cuore, Rome, Italy

²⁹ Polyposis Registry, St. Mark's Hospital, London, UK

³⁰ Inherited Tumour Syndromes Research Group, Institute of Cancer & Genetics, Cardiff University, UK

³¹ National Centre for Colorectal Disease, St. Vincent's University Hospital, Dublin, Ireland

³² Department of Surgery and Cancer, Imperial College, London UK

³³ Huntsman Cancer Institute, University of Utah, Salt Lake City, Utah, USA

³⁴ Department of Oncological Sciences, School of Medicine, University of Utah, Salt Lake City, Utah, USA

³⁵ Centre for Human Genetics, University of Marburg, Germany.

³⁶ National Center for Hereditary Tumor Syndromes, University Hospital Bonn, Bonn, Germany

Corresponding Author

Prof. Dr. Stefan Aretz, MD

Institute of Human Genetics

Center for Hereditary Tumor Syndromes

Biomedical Center

University Hospital Bonn

Venusberg-Campus 1

D-53127 Bonn

Tel.: + 49 (0) 228 - 287 51009

Fax: + 49 (0) 228 - 287 51011

ABSTRACT

Background: Pathogenic constitutional *APC* variants underlie familial adenomatous polyposis, the most common hereditary gastrointestinal polyposis syndrome. To improve variant classification and resolve the interpretative challenges of variants of uncertain significance (VUS), *APC*-specific ACMG/AMP variant classification criteria were developed by the ClinGen-InSiGHT Hereditary Colorectal Cancer/Polyposis Variant Curation Expert Panel (VCEP).

Methods: A streamlined algorithm using the *APC*-specific criteria was developed and applied to assess all *APC* variants in ClinVar and the InSiGHT international reference *APC* LOVD variant database.

Results: A total of 10,228 unique *APC* variants were analysed. Among the ClinVar and LOVD variants with an initial classification of (Likely) Benign or (Likely) Pathogenic, 94% and 96% remained in their original categories, respectively. In contrast, 41% ClinVar and 61% LOVD VUS were reclassified into clinically actionable classes, the vast majority as (Likely) Benign. The total number of VUS was reduced by 37%. In 21 out of 36 (58%) promising *APC* variants that remained VUS despite evidence for pathogenicity, a data mining-driven work-up allowed their reclassification as (Likely) Pathogenic.

Conclusions: The application of *APC*-specific criteria substantially reduced the number of VUS in ClinVar and LOVD. The study also demonstrated the feasibility of a systematic approach to variant classification in large datasets, which might serve as a generalisable model for other gene-/disease-specific variant interpretation initiatives. It also allowed for the prioritization of VUS that will benefit from in-depth evidence collection. This subset of *APC* variants was approved by the VCEP and made publicly available through ClinVar and LOVD for widespread clinical use.

Keywords: ACMG/AMP variant classification guidelines; *Adenomatous polyposis coli (APC)*; ClinGen; Familial adenomatous polyposis (FAP); InSiGHT

1 INTRODUCTION

2 Familial adenomatous polyposis (FAP, OMIM #175100) is an autosomal dominant precancerous
3 condition and the most common monogenic gastrointestinal polyposis syndrome, caused by
4 constitutional (germline) pathogenic variants (PV) in the tumour suppressor gene *APC* (1-3). The
5 colorectal phenotype exhibits high inter- and intra-familial variability from the growth of less than
6 100 up to thousands of adenomatous polyps (4). Surveillance colonoscopy and/or prophylactic
7 (procto)colectomy are offered to prevent colorectal cancer or delay disease progression (5-7).
8 Hence, the identification of an *APC* PV has direct relevance for the patients and their relatives,
9 defining *APC* as a highly clinically actionable gene (8). *APC* is a large gene composed of 15 coding
10 exons which encodes a multifunctional protein with several functional domains. Depending on the
11 colorectal phenotype and family history, causative *APC* variants can be identified in up to 85% of
12 patients with adenomatous polyposis (9-13), the vast majority of which are nonsense and frameshift
13 variants leading to a truncated protein with abrogated function (5-7, 14). During the last three
14 decades, thousands of rare *APC* PV have been identified in patients with FAP. Variants are
15 distributed across the gene, the majority of which are private, observed in only one or very few
16 families (www.lovd.nl/APC).

17 On the other hand, advances in high-throughput sequencing with the widespread implementation of
18 large multi-gene panel testing (MGPT) and exome/genome sequencing have generated an additional
19 plethora of *APC* variants in (healthy) individuals without a polyposis phenotype, many of which are
20 missense alterations. In the absence of comprehensive data and consensus for the level of evidence
21 required to corroborate variant interpretation, most of these variants remain Variants of Uncertain
22 Significance (VUS) or variants with conflicting assertions, accounting for around 67% of *APC* variants
23 in ClinVar. These VUS confer medical uncertainty and pose challenges in the clinical management of
24 patients and their relatives.

1 Since its inception, the ACMG/AMP guideline has evolved through further refinements to the various
2 variant assessment methods and evidence codes (15-21) and the development of gene- or disease-
3 specific ACMG/AMP classification criteria by variant curation expert panels (VCEPs) under the
4 governance of ClinGen (Clinical Genome Resource) (22). The International Society for
5 Gastrointestinal Hereditary Tumours (InSiGHT) is a research consortium that houses and curates the
6 world's largest databases for variants of gastrointestinal cancer-predisposing genes in the Leiden
7 Open Variation Database (LOVD) (23). Recently, a ClinGen-InSiGHT Hereditary Colorectal Cancer/
8 Polyposis VCEP (HCCP VCEP) was established (<https://clinicalgenome.org/affiliation/50099/>). The
9 APC subcommittee (APC VCEP) developed and validated APC-specific ACMG/AMP classification
10 criteria (24), readying the VCEP for variant submissions to ClinVar as an FDA Recognized Expert Panel.
11 The most updated version of the VCEP specifications can be found at
12 <https://cspec.genome.network/cspec/ui/svi/doc/GN089>.

13 In this study, we used the APC-specific criteria to perform a large-scale reclassification exercise of all
14 APC variants listed in ClinVar and the InSiGHT APC reference database LOVD. The criteria were
15 embedded and applied in a streamlined algorithm, which was supplemented by further data mining
16 and curation to achieve the most accurate classification. The results were compared to their original
17 classifications in respective databases and any discrepancies or clinical conflicts were addressed.

18 **METHODS**

19 **APC variant database merging and centralisation**

20 Prior to the extraction of variants, the landscape of all publicly available databases containing APC
21 variants was identified and examined for activity and curation status. Of at least 19 APC databases,
22 nine are inactive and another three are not curated (Supplementary Table 1). All listed curators, in
23 particular those of inactive, outdated, or orphaned databases were contacted to request sharing and
24 merging of data with the reference LOVD (v.3.0) installation. To establish a centralised, curated data
25 source of APC variants with consistent reporting format and phenotypic description, the InSiGHT APC

26 LOVD (<https://www.insight-database.org/genes/APC>) and the Global Variome shared LOVD
27 (www.lovd.nl/APC) were subsequently merged to generate one international reference *APC* gene
28 variant database in LOVD, abbreviated in the following as LOVD and accessible via all three URLs.

29 **ClinVar and LOVD variant extraction and annotation**

30 ClinVar variants with summary evaluation and individual submitter annotations were retrieved from
31 the March 2022 XML file. All alleles associated with the *APC* gene were extracted. The merged *APC*
32 LOVD database was downloaded on 12/05/2022. The legacy description of published variants was
33 recorded alongside their standardised nomenclature as per the HGVS (Human Genome Variation
34 Society) guidelines on the preferred reference transcript NM_000038.6 (25), correcting for any
35 errors where possible. All non-structural variants were annotated using the Ensembl Variant Effect
36 Predictor (VEP) (26). Structural variants defined by genomic alterations greater than 50bp in size
37 (gross deletions, duplications, inversions, in-frame, Alu and SVA retrotransposon insertions,
38 inversions and complex variants) were annotated manually using Mutalyzer (27).

39 **Reclassification algorithm**

40 Details of the *APC*-specific criteria were as published previously (24). A stepwise algorithm
41 encompassing all evidence codes was designed to systematically evaluate all *APC* variants in ClinVar
42 and LOVD (Figure 1).

43 Minor allele frequency data (BA1, BS1, PM2_supporting)

44 The frequency of all *APC* variants in reference populations were compared against the minor allele
45 frequency (MAF) criteria BA1 (≥ 0.001), BS1 (≥ 0.00001), and PM2_supporting (≤ 0.000003 or absent;
46 first version of criteria). The non-cancer datasets from gnomAD (the Genome Aggregation Database)
47 v2.1.1 and v3.1.2 were used as the reference population frequency data for non-structural variants
48 (28). Exome sequencing data from 323,228 healthy individuals without a diagnosis of CRC in the UK
49 Biobank were also used to further enhance the detection of rare *APC* variants (29). If a variant was
50 present in multiple reference population datasets, the highest MAF was calculated from any

51 subpopulation with more than 2000 alleles, with the exclusion of founder populations. The
52 frequency of structural variants was examined in gnomAD SVs v2.1 and the DGV (Database of
53 Genomic Variants) Gold Standard release from 15/05/2016 (30).

54 Predictive data (PVS1, PP3, BP1, BP4, BP7)

55 Truncating variants, canonical \pm 1/2 splice site variants, and exonic last nucleotide guanine to non-
56 guanine variants were assigned the Loss-of-Function (LoF) criterion PVS1 if they are located between
57 codon 49 and 2645 inclusive (24). Splice prediction was performed using SpliceAI and MaxEntScan
58 via VEP, which determined PP3 and BP4 eligibility for synonymous and intronic variants and PP3
59 eligibility for presumed missense variants to reveal possible splicing effects (31, 32). Variants
60 exceeding a score of 0.6 in SpliceAI and a 15% reduction from the native site prediction in
61 MaxEntScan were considered spliceogenic, and variants with a SpliceAI score of less than 0.2 and a
62 MaxEntScan score of less than 3 were considered to have no impact on splicing. The criteria were
63 only applied when both prediction tools showed concordant results. BP7 was subsequently applied
64 to synonymous and deep intronic variants at or beyond +7/-21 which satisfied BP4. The missense
65 code BP1 was applied to missense variants located outside of the first 15-amino acid repeat of the β -
66 catenin binding domain (codon 1021-1035) if they were consistently deemed non-spliceogenic by
67 SpliceAI and MaxEntScan.

68 Experimental data (BS3, PS3)

69 Published mRNA splicing assays and protein function assays of *APC* variants were collated in a
70 systematic review by the *APC* VCEP to derive gene-specific recommendations for the application of
71 the experimental criteria BS3 and PS3 (24). VCEP-approved experimental evidence for *APC* variant
72 classifications included RNA assays for PS3 and BS3, β -catenin regulated transcriptional assays for
73 PS3_supporting and BS3_supporting and surface plasmon resonance assays for PS3_supporting. The
74 proportion of aberrant transcripts, evidence of biallelic expression and use of nonsense-mediated

75 decay inhibition were also noted from the original publications when available, which determined
76 the quality of the data and therefore the weight assigned for PS3 and BS3.

77 Clinical data (PS4, PS2, PM6, PP1, BS4, BP2, BP5, BS2)

78 The phenotype details of individuals with *APC* variants from the InSiGHT LOVD download were
79 retrieved. In cases where phenotype details were recorded as unstructured text, text mining was
80 employed to extract and stratify useful information. Affected individuals were scored for PS4 using
81 methods as described previously (24). Further data mining was undertaken focusing on the
82 identification of confirmed *de novo APC* variants (PS2/PM6), segregation and non-segregation
83 analysis (PP1/BS4), co-occurrence of the variant under assessment with other established
84 pathogenic *APC* variant (BP2) or with an alternative molecular basis of disease (BP5) including
85 heterozygous PV in *POLD1* or *POLE* (Polymerase-proofreading-associated polyposis), and biallelic PV
86 either in *MUTYH* (*MUTYH*-associated polyposis), in *NTHL1* (*NTHL1*-associated tumour syndrome), in
87 *MSH3* (*MSH3*-associated polyposis), or the MMR genes *MLH1*, *MSH2*, *MSH6*, or *PMS2* (Constitutional
88 MMR deficiency). The Human Gene Mutation Database (HGMD) and Universal Mutation Database
89 (UMD) were also examined for additional data (33, 34). For variants that were already listed in
90 gnomAD, their presence in healthy unaffected individuals in the UK Biobank non-CRC population was
91 regarded as additional evidence in support of a benign classification (BS2). The absolute number of
92 heterozygous individuals in the UK Biobank non-CRC dataset was counted, scored according to the
93 *APC* VCEP's definition for a healthy individual, and given the healthy control code BS2 with
94 appropriate weight. As homozygous LoF *APC* variants were shown to be lethal at embryonic
95 developmental stages (35), the observation of variant homozygosity ≥ 2 times in any reference
96 population database was also considered strong evidence for a benign classification as defined in the
97 *APC*-specific criteria.

98 **Review and synthesis of primary criteria combination**

99 The *APC*-specific criteria applied for each variant were pulled together to generate a preliminary
100 criteria combination according to the *APC* VCEP's rules for combining criteria (24). In addition to the
101 phenotype description in LOVD, for variants that fulfilled a reasonable set of predictive and/or
102 experimental criteria (e.g., PVS1, PS3 or BS3) but remained unclassified, the corresponding internal
103 clinical records and/or reference publication of the variant in question was consulted to curate
104 additional phenotype points, and further data mining was conducted using Mastermind, LitVar and
105 PubMed as required to reach a non-VUS classification. Additionally, the classification of the 58
106 published *APC* variants in the pilot study for the development of the *APC*-specific guideline were
107 incorporated, adding another layer of data and expert review to the result. Finally, the revised
108 criteria combination was used to calculate a final pathogenicity class. Missense variants at the same
109 codon and spliceogenic variants at the same nucleotide position as established P or LP variants were
110 identified and the criteria PS1 or PM5 were applied accordingly.

111 **Prioritised list of variants for further review**

112 The revised variant classification was compared with the original pathogenicity assertion in ClinVar
113 and LOVD. Variants with clinically significant conflict (B/LB vs. P/LP) between the original assertion
114 and reclassification were examined for causes of discrepancy. From the variants that remained VUS
115 after reclassification, 36 variants close to reaching a pathogenic classification were identified (Table
116 1 and Figure 2), which included (1) truncating VUS that fulfilled PVS1 but not PM2_supporting; (2)
117 VUS with pathogenic *in silico* predictions; (3) VUS with experimental findings suggestive of
118 deleterious effect and (4) VUS observed in individuals with FAP-associated phenotypes. A targeted
119 literature search was conducted to acquire further information for all 36 variants. If a clinically
120 relevant classification could be achieved, no further work-up was done (group 1 in Table 1 and
121 Figure 2). If this was not the case, phenotype data were requested for these variants from VCEP
122 members and internal laboratory contributors in a standardised clinical data request form and
123 analysed with the aim to resolve conflicting criteria and upgrade classification (group 2 in Table 1
124 and Figure 2).

125 RESULTS

126 Variant assessment using the reclassification algorithm based on VCEP APC-specific criteria

127 A total of 10,228 APC variants were analysed in this study, which included 190 (2%) structural and
128 10,038 (98%) non-structural variants (Supplementary Table 2). A total of 9,121 and 1,887 APC
129 variants were present in ClinVar and LOVD, respectively (Figure 3); 780 (41% of LOVD) variants are
130 shared between ClinVar and LOVD. The largest group in the ClinVar dataset is the VUS class (67%),
131 followed by the B/LB group (20%) and P/LP group (14%). Most variants in LOVD are P/LP (76%),
132 whereas VUS only account for 16% and B/LB variants for 8%.

133 Comparison between the original ClinVar or LOVD assertions and the revised classification are shown
134 in Figure 3a. By applying the APC-specific criteria, in the ClinVar dataset, 4,869 variants were
135 classified as B/LB (53%), 1,185 as P/LP (13%), and 3,787 as VUS (42%). In the LOVD dataset, 299
136 variants were classified as B/LB (16%), 1,406 as P/LP (75%), and 182 as VUS (10%). Notably, 94% of
137 ClinVar variants with an initial classification of B/LB or P/LP remained in their respective benign and
138 pathogenic categories after reclassification. Similarly, 96% LOVD variants were reclassified without
139 significant clinical conflict. A considerable portion of previous P variants were downgraded to LP,
140 which included 768 out of 1034 (74%) and 920 out of 1,368 (67%) previous P variants in ClinVar and
141 LOVD, respectively. Remarkably, 40% of ClinVar VUS were reclassified as B/LB and 1% as P/LP, while
142 in LOVD 55% VUS were reclassified as B/LB and 6% as P/LP. The percentage of VUS was reduced
143 from 67% to 42% in ClinVar and from 16% to 10% in LOVD. As a whole, the total number of VUS in
144 the combined dataset was reduced by 37% from 6,142 to 3,865 [38% in ClinVar (from 6,074 to 3,787)
145 and 40% in LOVD (from 304 to 182)].

146 Only one previously LP variant was reclassified as LB by the APC-specific criteria: c.70C>T
147 (p.Arg24Ter). Since this variant is 5' of codon 49, PVS1 is not applicable. This variant was present in 6
148 out of 268,098 alleles with a MAF of 0.00002 in the gnomAD v.2.1.1 non-cancer population, meeting
149 the initial BS1 specification (version 1.0.0). Meanwhile in the most recent version of the criteria

150 (version 2.1.0), the APC VCEP has recommended the use of the “filtering allele frequency” (FAF) for
151 BA1 and BS1, this value is not available for this variant (BS1 not met). Based on a survey among the
152 VCEP members and a literature search (36, 37), 67 heterozygote carriers of c.70C>T were identified,
153 which are reported either without or with a very mild polyposis phenotype (only 3 carriers fulfilled
154 0.5 phenotype points each; PS4_supporting). There are also more than 10 unaffected healthy
155 individuals \geq 50 years in this survey, fulfilling BS2. As a result, the variant was classified as LB and
156 PS4_supporting was not considered as conflicting.

157 A number of previously B/LB or P/LP variants were reclassified as VUS, which included 73 B/LB (0.7%)
158 and 117 P/LP (1.1%) variants in ClinVar, and 13 B/LB (0.1%) and 51 P/LP (0.05%) variants in LOVD,
159 respectively. The reasons underlying these changes are outlined in Supplementary Table 3 and
160 explored in further detail in the discussion section.

161 The distribution of variants by type is shown in Figure 3b with the majority being missense variants
162 (42%), followed by synonymous or intronic variants (25%), and truncating (frameshift/nonsense)
163 variants (19%). The original (O) and revised (R) pathogenicity class based on the algorithm was
164 compared for each variant type. 97% of frameshift/nonsense variants are P/LP and 99% of
165 synonymous and intronic variants are B/LB according to the APC-specific criteria. In the original
166 ClinVar and LOVD assertions, 99% of putative missense variants are VUS. This proportion was
167 reduced significantly to 71% using the APC-specific criteria, where approximately 30% of all missense
168 variants were reclassified as B/LB. Application of the APC-specific criteria reduced the percentage of
169 synonymous/intronic (at or beyond +7/-21 intronic positions) VUS from 38% to 1%, classifying the
170 vast majority as B/LB. For variants flanking splice sites (within +7/-21 intronic positions), the range of
171 classifications was similar before and after reclassification. 31% in-frame variants and 42% UTR
172 variants were also classified as B/LB, while prior to reclassification they were mostly VUS. After
173 application of the specific criteria, 3,865 variants remained VUS, which included 3,067 missense
174 variants (79%), and low numbers of other variant types (Figure 3).

175 **Impact and usage of the APC-specific codes**

176 The frequency of use of each APC-specific code is noted in Supplementary Table 4. All criteria were
177 used at least once during the reclassification process. The most frequently applied codes were the
178 two pathogenic criteria PM2_supporting (used for 69% of all 10,228 variants) and PVS1 (21%), and
179 the four benign criteria BP1 (41%), BP4 (27%), BP7 (22%), and BS1 (20%), while on the other hand,
180 half of the codes are used for less than 1% of variants. 2,192 of all 2568 truncating variants (85%)
181 were assigned PVS1_variable (including frameshift, nonsense, spliceogenic, gross
182 deletions/duplications and last nucleotide of exon G to non-G variants). In contrast, PS4_variable
183 could only be applied in 455 of the 2,208 individuals (21%) with annotated phenotype description in
184 LOVD. 3,145 variants (31% of all variants) were present in the gnomAD v.2.1.1 non-cancer
185 population and/or the UK Biobank non-CRC dataset, of which 427 (14%; 4% of all variants) could be
186 classified as benign just by the BA1 code alone and 135 (4%; 1% of all variants) as likely benign just
187 by the BS1 code alone.

188 The most common criteria combination was BP1 and PM2_supporting: 2,622 variants (26%) were
189 missense variants that were absent from population databases and had a consistent benign splice
190 prediction, resulting in a VUS classification. The second most common criteria combination was PVS1
191 and PM2_supporting, which applied to 2,165 variants: 77% (1,667/2,165) of these variants (16% of
192 all variants) were classified as LP solely based on the two codes. This was followed by BP4 and BP7
193 leading to a LB classification for 1,708 synonymous and intronic variants (17% of all variants). The
194 combination of BS1 and BP1 was given to 1,064 missense variants (10% of all variants) which
195 resulted in a LB classification. 155 variants (1.5% of all variants) were present in the non-CRC control
196 dataset of UK Biobank, which fulfilled the definition for a healthy unaffected individual in the APC-
197 specific criteria and allowed the assessment for BS2_variable.

198 **Further data mining and criteria review for a prioritised list of variants**

199 After the initial application of the *APC*-specific criteria, a considerable fraction of VUS (63%)
200 remained as expected. We selected 36 promising variants from these remaining VUS with some
201 evidence for pathogenicity (details see Materials & Methods), which formed a prioritised list of
202 variants for further review as outlined in the workflow (Figure 2).

203 In 11 truncating variants (Table 1, encompassing both Group 1 and 2), the relegation of their prior
204 LP/P classification were due to their presence at very low frequencies in reference population
205 databases, in this case the occurrence of one allele in a gnomAD non-cancer subpopulation.
206 Depending on the denominator (i.e., size of the subpopulation), 7 previously LP/P truncating variants
207 were precluded from the use of PM2_supporting and 4 even fulfilled threshold for BS1 using version
208 1.0.0 of the *APC*-specific criteria. To resolve this issue, the *APC* VCEP added a caveat to
209 PM2_supporting in the criteria version 2.1.0 where the allele frequency threshold of $\leq 0.0003\%$
210 (0.000003) is only used if the allele count is > 1 . To tolerate singleton allele occurrence in gnomAD,
211 the *APC* VCEP set an allele frequency of $< 0.001\%$ (0.00001) (lower than BS1) if the allele count is ≤ 1 .
212 Moreover, the *APC* VCEP recommended in criteria version 2.1.0 the use of the FAF for BA1 and BS1
213 to avoid the issue of singleton alleles satisfying the allele frequency criteria. This allowed the use of
214 PM2_supporting and the reclassification of these 11 variants as LP/P as shown in Table 1.

215 For five variants that were originally P/LP in ClinVar or LOVD but reclassified as VUS by the algorithm,
216 a more extensive data mining and literature review led to the return of P as their final classification
217 (Group 1 in Table 1 and Figure 2). The previous pathogenic classification would indicate the
218 observation of these variants in affected individuals on multiple occasions and a targeted search was
219 finally informative.

220 For the remaining 31 variants (Group 2 in Table 1 and Figure 2), internal lab contributors provided
221 available phenotypic information that allowed the upgrade of classification from VUS to P/LP for 11
222 (35%) variants (Group 2 in Table 1 and Figure 2). Five variants of Group 2 were evaluated as LP based
223 on the reassessment of the MAF criteria, but no relevant phenotypic information was available.

224 Overall, further data mining of selected representative variants resulted in the enhanced
225 classification of 21 out of 36 variants (58%) into meaningful pathogenicity classes.

226 **DISCUSSION**

227 The rising number of VUS in clinically actionable genes such as *APC* represents an important issue in
228 the post-genomic era that hinders the translation of genetic diagnostics into improved health
229 outcomes. In this study, we first identified the current landscape of *APC* variants with respect to
230 their original assertions in the two most important international databases for *APC* variants: ClinVar
231 and the InSiGHT LOVD. A striking difference was noted in the distribution of pathogenicity classes:
232 while around two-thirds of ClinVar variants were originally VUS and 20% B/LB, roughly 70% of *APC*
233 variants on LOVD were P/LP (Figure 3). This is not unexpected since variants submitted to LOVD are
234 usually detected in patients with the relevant phenotype (i.e., clinically evident colorectal
235 adenomatous polyposis) where the detection of PV is more likely and the primary focus of curators
236 of locus-specific databases. Conversely, data in ClinVar is derived from the increased application of
237 high-throughput sequencing methods in patients with less specific or unrelated phenotypes and
238 healthy individuals has generated an extensive catalogue of rare variants in actionable genes, the
239 majority of which is expected to be benign or have a low penetrance. However, in the absence of
240 overwhelming evidence, they are usually conservatively classified as VUS.

241 This study was the first application of the full set of ClinGen-approved gene-specific ACMG/AMP
242 guidelines (24) to a large set of variants identified worldwide. We set out to reclassify as many *APC*
243 variants as possible with the aim to improve consistency and accuracy in *APC* variant classification.
244 By consolidating *APC* variants in ClinVar and LOVD, we provide a holistic overview of most *APC*
245 variants identified to date using gene-specific criteria in individuals with a spectrum of diseases.
246 Other smaller scale studies used gene-/disease-specifications that covered only partial evidence
247 domains (38), or meta-classification methods such as the multifactorial likelihood analysis (39, 40).

248 One of the major findings of this study is that the application of the *APC*-specific criteria reduced the
249 number of VUS by 37% collectively in ClinVar and LOVD such that 2,277 variants could be revised
250 into clinically meaningful classes. The majority (40%) of the former VUS were reclassified as B/LB (n=
251 2441) owing to their presence in reference population databases fulfilling MAF criteria (BA1/BS1)
252 and/or computational data predicting no impact on protein function (BP1/BP4/BP7). As a result,
253 these B/LB variants no longer need to be considered in the diagnosis of FAP and reported in genetic
254 testing, which in turn minimises the anxiety and potential overtreatment in affected individuals
255 worldwide (41). On the other hand, 51 previous VUS were reclassified as P/LP (0.8%), a result which
256 confirms the genetic diagnosis of FAP and enables the timely management and predictive testing of
257 all at-risk relatives. On the other hand, 95% and 94% of the previously B/LB and P/LP variants
258 remained in the B/LB and P/LP classifications, respectively. These findings demonstrate that the
259 application of the *APC*-specific criteria is highly effective in substantially improving the
260 reclassification of VUS into clinically relevant pathogenicity classes while preserving the original
261 interpretation of variants with existing evidence-based classifications.

262 The only variant [c.70C>T;(p.Arg24Ter)] with a clinically significant change from LP to LB was due to
263 the indiscriminate use of PVS1 by the original submitters. Historically, variants are generally assumed
264 to be pathogenic if they are protein truncating, without considering the impact of alternative start
265 codon, alternative splicing, the preservation of relevant functional domains, and nonsense mediated
266 decay. The *APC* VCEP conservatively defined that PVS1 is only applicable for truncating variants
267 between codon 49 and 2645 inclusive (24). The *APC* variant c.70C>T;(p.Arg24Ter) is 5' of that cut-off
268 and has sufficient clinical evidence to suggest a LB classification as discussed in the results.

269 In the combined databases, a total of 128 P/LP variants (92 non-structural, 37 structural) and a total
270 of 87 B/LB variants (86 structural, 1 non-structural) were reclassified as VUS (216 in total, 5% of
271 previous non-VUS), which is clinically significant for the previous P/LP variants and will affect the
272 diagnosis, counselling, and the predictive power of genetic testing (Supplementary Table 3). Among

273 the 128 P/LP-to-VUS variants, 14 were gross deletions, which affected only the promoter region
274 and/or exon 2 of the NM_000038.6 transcript (i.e., the first coding exon 1). The exact molecular
275 consequences are difficult to ascertain, and they could not be properly classified in the absence of
276 convincing clinical and segregation data although VCEP members thought they were likely
277 pathogenic. Another 21 variants were gross duplications which had unknown or only presumed
278 impact on the reading frame and lacking clinical data. While many *APC* gross duplications are indeed
279 located in a tandem position (4, 42, 43), their impact on the transcript is unknown per se, especially
280 if they extended beyond the open reading frame of the gene. Seven 5'UTR variants, nine truncating
281 variants at the 5' end and 7 truncating variants at the 3' end of the gene were excluded from the
282 application of PVS1 and were classified as VUS.

283 In addition to varying PVS1 application for frameshift and nonsense variants, we also noted that
284 canonical $\pm 1/2$ splice site changes and intronic variants flanking the splice sites (+7/-21 bp) were
285 predominantly assessed as pathogenic by the submitters to LOVD and ClinVar, although splice
286 predictions and transcript analyses might have suggested otherwise (i.e., weak native site, leaky
287 splicing control etc). Among the 62 intronic variants in the vicinity of splice sites that were
288 reclassified as VUS, 43 (69%) were originally LP/P and 19 (30%) were originally LB/B. The remaining
289 variants included 55 deep intronic, synonymous, in-frame and UTR variants and 25 missense variants,
290 where the reclassification as VUS was the result of a combination of scarcity of clinical data and non-
291 contributory *in silico* predictions. These variants were better evaluated by the cautionary,
292 conservative assessment intrinsic to the *APC*-specific criteria and highlights the importance of
293 curating evidence from different domains.

294 In the *APC*-specific criteria, a range of evidence weight adjustments is specified as means of improve
295 precision and quality. The lack of detailed evidence description in many studies meant that certain
296 *APC*-specific criteria can only be applied at lower weights than what the submitters of ClinVar and
297 LOVD might have intended. In this study, clinical data was only extracted from individuals with

298 phenotype data in LOVD and for a selected number of variants where clinical data was essential for
299 their reclassification into a non-VUS category. ClinVar and LOVD submitters may have additional
300 variant information not apparent to this data-mining process. This is confirmed in the further
301 curation of selected variants, where 56% of these VUS could again be reclassified into P/LP by
302 additional data mining (Table 1). A key step was the request of further clinical information from our
303 internal laboratory contributors, which substantially improved the interpretation of variants on the
304 verge of being classified as LP/P.

305 A considerable number of previous P variants (75% in ClinVar and 68% in LOVD) were downgraded
306 to LP (e.g. truncating variants fulfilling only PVS1 and PM2_supporting). This phenomenon is also
307 known from other genes where gene-specific rules are being applied. Submitters may be overrating
308 the quality of available pathogenic evidence, have additional evidence, applying the original
309 moderate strength for PM2 or weighing the clinical information differently based on internal
310 calibrations. A large-scale reclassification project is not able to screen for all available evidence for
311 thousands of variants due to limited time and personal resources. Consequently, it cannot be
312 excluded that some reclassified LP variants are in fact P, which has a great potential for future
313 upgrade through diligent reporting of clinical information and data-sharing. In practical terms a LP
314 classification has a posterior probability of pathogenicity of 0.9 to 0.99, which nonetheless demands
315 clinical action when detected (44, 45).

316 After reclassification, 80% of the remaining VUS were presumed missense variants. The fundamental
317 mechanism of *APC* pathogenicity is based on the loss of a large 3' part of the protein which includes
318 the relevant functional domains. In addition to further evidence of functional redundancy in the *APC*
319 protein, the central and C-terminal domains of the *APC* protein are natively unfolded, all of which
320 likely explains the resistance of the *APC* protein to missense variation (46) and the absolute
321 predominance of truncating alterations among *APC* PV. The vast majority of the remaining missense
322 *APC* VUS in ClinVar are likely benign incidental findings from non-targeted testing in individuals with

323 unrelated phenotypes, for which *in silico* prediction tools and (massively parallel) functional assays
324 are unlikely to contribute significantly to improve classifications. The recruitment of variant data
325 from large reference population projects such as the UK Biobank on the other hand, is the key to
326 determine the pathogenicity of these missense variants.

327 This reclassification endeavour was also considered a proof-of-concept study for the ongoing
328 method of operation of VCEPs (Figure 2). The review and discussion of every single submitted
329 constitutional variant in the *APC* gene is unrealistic with respect to the resources currently provided
330 to a VCEP, until appropriate bioinformatic tools become available in the future. In this study, we
331 developed a stratified variant curation process, whereby the relatively straightforward variants could
332 be processed in batch and become candidates for fast-track VCEP review (e.g. variants fulfilling BA1
333 or BS1 plus BP1 without any conflicting classifications submitted to ClinVar or LOVD) in an updated
334 variant curation interface (VCI) to streamline the variant approval process. We identified several
335 prioritised groups of variants which can be the subject of further targeted literature review (Suppl
336 Table 3), data-mining, and clinical data request from database submitters and internal lab
337 contributors, a process to enhance variant interpretation as demonstrated in this study. These
338 challenging variants can also form the basis for scientific follow-up studies to evaluate the causal
339 relationships using additional investigations such as segregation or transcript analyses.

340 To our knowledge, this is the first study to undertake a large-scale approach to the interpretation of
341 VUS in ClinVar and LOVD using ClinGen-approved gene-specific criteria. We developed and applied
342 an algorithm for variant reclassification and demonstrated that the application of *APC*-specific
343 criteria can substantially alleviate the burden of VUS, thereby laying the groundwork for a
344 prospective streamlined expert panel approval of clinically actionable *APC* variants in the VCI. By
345 using the VCEP specifications, diagnostic laboratories could reduce their rates of reporting VUS.

346 This study highlights the utility of a systematic, data-driven analysis using gene-specific ACMG/AMP
347 criteria, complemented by further targeted data-mining and clinical data requests. By this approach,

348 this study marks the initiation of a dynamic, long-term curation process for the *APC* VCEP. The
349 suggested workflow also serves as a generalisable model of operation for other gene-/disease-
350 specific variant interpretation initiatives, achieving accurate and highly efficient variant
351 interpretation based on an array of carefully curated evidence. To further improve VUS
352 interpretation and provide clinically informative variant classification beyond this approach, the
353 availability of more population-based datasets, and user-friendly modes of sharing clinical and
354 molecular data are needed; a challenge that has to be handled by the respective expert communities
355 and data submitters.

Funding

This publication was supported in part by the National Human Genome Research Institute of the National Institutes of Health for the Baylor College of Medicine/Stanford University Clinical Genome Resource-2U24HG009649 and from the National Cancer Institute U24 Curation Panels through the U24CA258119. It also was supported in part by the Intramural Research Program of the National Library of Medicine, National Institutes of Health.

Acknowledgements

This work was supported (not financially) by the European Reference Network on Genetic Tumour Risk Syndromes (ERN GENTURIS)- Project ID No 739547. ERN GENTURIS is partly co-funded by the European Union within the framework of the Third Health Programme “ERN-2016—Framework Partnership Agreement 2017-2021.”

References

1. Kinzler KW, Nilbert MC, Su LK, Vogelstein B, Bryan TM, Levy DB, et al. Identification of FAP locus genes from chromosome 5q21. *Science*. 1991;253(5020):661-5.
2. Groden J, Thliveris A, Samowitz W, Carlson M, Gelbert L, Albertsen H, et al. Identification and characterization of the familial adenomatous polyposis coli gene. *Cell*. 1991;66(3):589-600.
3. WHO. The Introduction to Genetic Tumour Syndromes In: Board WCoTE, editor. *Genetic Tumour Syndromes 5th ed*2024.
4. Jasperson KW, Tuohy TM, Neklason DW, Burt RW. Hereditary and familial colon cancer. *Gastroenterology*. 2010;138(6):2044-58.
5. Syngal S, Brand RE, Church JM, Giardiello FM, Hampel HL, Burt RW. ACG clinical guideline: Genetic testing and management of hereditary gastrointestinal cancer syndromes. *Am J Gastroenterol*. 2015;110(2):223-62; quiz 63.
6. Monahan KJ, Bradshaw N, Dolwani S, Desouza B, Dunlop MG, East JE, et al. Guidelines for the management of hereditary colorectal cancer from the British Society of Gastroenterology (BSG)/Association of Coloproctology of Great Britain and Ireland (ACPGBI)/United Kingdom Cancer Genetics Group (UKCGG). *Gut*. 2020;69(3):411-44.
7. van Leerdam ME, Roos VH, van Hooft JE, Dekker E, Jover R, Kaminski MF, et al. Endoscopic management of polyposis syndromes: European Society of Gastrointestinal Endoscopy (ESGE) Guideline. *Endoscopy*. 2019;51(9):877-95.
8. Miller DT, Lee K, Abul-Husn NS, Amendola LM, Brothers K, Chung WK, et al. ACMG SF v3.1 list for reporting of secondary findings in clinical exome and genome sequencing: A policy statement of the American College of Medical Genetics and Genomics (ACMG). *Genetics in Medicine*. 2022;24(7):1407-14.
9. van der Luijt RB, Khan PM, Vasen HF, Tops CM, van Leeuwen-Cornelisse IS, Wijnen JT, et al. Molecular analysis of the APC gene in 105 Dutch kindreds with familial adenomatous polyposis: 67 germline mutations identified by DGGE, PTT, and southern analysis. *Hum Mutat*. 1997;9(1):7-16.

10. Wallis YL, Morton DG, McKeown CM, Macdonald F. Molecular analysis of the APC gene in 205 families: extended genotype-phenotype correlations in FAP and evidence for the role of APC amino acid changes in colorectal cancer predisposition. *J Med Genet.* 1999;36(1):14-20.
11. Friedl W, Caspari R, Sengteller M, Uhlhaas S, Lamberti C, Jungck M, et al. Can APC mutation analysis contribute to therapeutic decisions in familial adenomatous polyposis? Experience from 680 FAP families. *Gut.* 2001;48(4):515-21.
12. Bisgaard ML, Ripa R, Knudsen AL, Bülow S. Familial adenomatous polyposis patients without an identified APC germline mutation have a severe phenotype. *Gut.* 2004;53(2):266-70.
13. Rivera B, González S, Sánchez-Tomé E, Blanco I, Mercadillo F, Letón R, et al. Clinical and genetic characterization of classical forms of familial adenomatous polyposis: a Spanish population study. *Ann Oncol.* 2011;22(4):903-9.
14. Vasen HF, Mösllein G, Alonso A, Aretz S, Bernstein I, Bertario L, et al. Guidelines for the clinical management of familial adenomatous polyposis (FAP). *Gut.* 2008;57(5):704-13.
15. Jarvik GP, Browning BL. Consideration of Cosegregation in the Pathogenicity Classification of Genomic Variants. *Am J Hum Genet.* 2016;98(6):1077-81.
16. Abou Tayoun AN, Pesaran T, DiStefano MT, Oza A, Rehm HL, Biesecker LG, et al. Recommendations for interpreting the loss of function PVS1 ACMG/AMP variant criterion. *Hum Mutat.* 2018;39(11):1517-24.
17. Biesecker LG, Harrison SM. The ACMG/AMP reputable source criteria for the interpretation of sequence variants. *Genet Med.* 2018;20(12):1687-8.
18. Ghosh R, Harrison SM, Rehm HL, Plon SE, Biesecker LG. Updated recommendation for the benign stand-alone ACMG/AMP criterion. *Hum Mutat.* 2018;39(11):1525-30.
19. Brnich SE, Abou Tayoun AN, Couch FJ, Cutting GR, Greenblatt MS, Heinen CD, et al. Recommendations for application of the functional evidence PS3/BS3 criterion using the ACMG/AMP sequence variant interpretation framework. *Genome Med.* 2019;12(1):3.
20. Rofes P, Menéndez M, González S, Tornero E, Gómez C, Vargas-Parra G, et al. Improving Genetic Testing in Hereditary Cancer by RNA Analysis: Tools to Prioritize Splicing Studies and Challenges in Applying American College of Medical Genetics and Genomics Guidelines. *J Mol Diagn.* 2020;22(12):1453-68.
21. Smirnov D, Schlieben LD, Peymani F, Berutti R, Prokisch H. Guidelines for clinical interpretation of variant pathogenicity using RNA phenotypes. *Hum Mutat.* 2022;43(8):1056-70.
22. Rivera-Muñoz EA, Milko LV, Harrison SM, Azzariti DR, Kurtz CL, Lee K, et al. ClinGen Variant Curation Expert Panel experiences and standardized processes for disease and gene-level specification of the ACMG/AMP guidelines for sequence variant interpretation. *Hum Mutat.* 2018;39(11):1614-22.
23. Plazzer JP, Sijmons RH, Woods MO, Peltomäki P, Thompson B, Den Dunnen JT, et al. The InSiGHT database: utilizing 100 years of insights into Lynch syndrome. *Fam Cancer.* 2013;12(2):175-80.
24. Spier I, Yin X, Richardson M, Pineda M, Laner A, Ritter D, et al. Gene-specific ACMG/AMP classification criteria for germline APC variants: Recommendations from the ClinGen InSiGHT Hereditary Colorectal Cancer/Polyposis Variant Curation Expert Panel. *Genet Med.* 2024;26(2):100992.
25. den Dunnen JT, Dagleish R, Maglott DR, Hart RK, Greenblatt MS, McGowan-Jordan J, et al. HGVS Recommendations for the Description of Sequence Variants: 2016 Update. *Hum Mutat.* 2016;37(6):564-9.
26. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, et al. The Ensembl Variant Effect Predictor. *Genome Biol.* 2016;17(1):122.
27. Lefter M, Vis JK, Vermaat M, den Dunnen JT, Taschner PEM, Laros JFJ. Mutalyzer 2: next generation HGVS nomenclature checker. *Bioinformatics.* 2021;37(18):2811-7.
28. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alfoldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature.* 2020;581(7809):434-43.

29. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* 2015;12(3):e1001779.
30. MacDonald JR, Ziman R, Yuen RK, Feuk L, Scherer SW. The Database of Genomic Variants: a curated collection of structural variation in the human genome. *Nucleic Acids Res.* 2014;42(Database issue):D986-92.
31. Jaganathan K, Kyriazopoulou Panagiotopoulou S, McRae JF, Darbandi SF, Knowles D, Li Yi, et al. Predicting Splicing from Primary Sequence with Deep Learning. *Cell.* 2019;176(3):535-48.e24.
32. Yeo G, Burge CB. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol.* 2004;11(2-3):377-94.
33. Stenson PD, Ball EV, Mort M, Phillips AD, Shiel JA, Thomas NS, et al. Human Gene Mutation Database (HGMD): 2003 update. *Hum Mutat.* 2003;21(6):577-81.
34. Bérout C, Collod-Bérout G, Boileau C, Soussi T, Junien C. UMD (Universal mutation database): a generic software to build and analyze locus-specific databases. *Hum Mutat.* 2000;15(1):86-94.
35. Moser AR, Luongo C, Gould KA, McNeley MK, Shoemaker AR, Dove WF. ApcMin: a mouse model for intestinal and mammary tumorigenesis. *Eur J Cancer.* 1995;31a(7-8):1061-4.
36. Yurgelun MB, Kulke MH, Fuchs CS, Allen BA, Uno H, Hornick JL, et al. Cancer Susceptibility Gene Mutations in Individuals With Colorectal Cancer. *J Clin Oncol.* 2017;35(10):1086-95.
37. Kanter-Smoler G, Fritzell K, Rohlin A, Engwall Y, Hallberg B, Bergman A, et al. Clinical characterization and the mutation spectrum in Swedish adenomatous polyposis families. *BMC Med.* 2008;6:10.
38. Denham NC, Pearman CM, Ding WY, Waktare J, Gupta D, Snowdon R, et al. Systematic re-evaluation of SCN5A variants associated with Brugada syndrome. *J Cardiovasc Electrophysiol.* 2019;30(1):118-27.
39. Thompson BA, Spurdle AB, Plazzer JP, Greenblatt MS, Akagi K, Al-Mulla F, et al. Application of a 5-tiered scheme for standardized classification of 2,360 unique mismatch repair gene variants in the InSiGHT locus-specific database. *Nat Genet.* 2014;46(2):107-15.
40. Parsons MT, Tudini E, Li H, Hahnen E, Wappenschmidt B, Feliubadaló L, et al. Large scale multifactorial likelihood quantitative analysis of BRCA1 and BRCA2 variants: An ENIGMA resource to support clinical variant classification. *Human Mutation.* 2019;40(9):1557-78.
41. Welsh JL, Hoskin TL, Day CN, Thomas AS, Cogswell JA, Couch FJ, et al. Clinical Decision-Making in Patients with Variant of Uncertain Significance in BRCA1 or BRCA2 Genes. *Ann Surg Oncol.* 2017;24(10):3067-72.
42. Kaufmann A, Vogt S, Uhlhaas S, Stienen D, Kurth I, Hameister H, et al. Analysis of rare APC variants at the mRNA level: six pathogenic mutations and literature review. *J Mol Diagn.* 2009;11(2):131-9.
43. McCart A, Latchford A, Volikos E, Rowan A, Tomlinson I, Silver A. A novel exon duplication event leading to a truncating germ-line mutation of the APC gene in a familial adenomatous polyposis family. *Fam Cancer.* 2006;5(2):205-8.
44. Plon SE, Eccles DM, Easton D, Foulkes WD, Genuardi M, Greenblatt MS, et al. Sequence variant classification and reporting: recommendations for improving the interpretation of cancer susceptibility genetic test results. *Hum Mutat.* 2008;29(11):1282-91.
45. Tavtigian SV, Greenblatt MS, Harrison SM, Nussbaum RL, Prabhu SA, Boucher KM, et al. Modeling the ACMG/AMP variant classification guidelines as a Bayesian classification framework. *Genet Med.* 2018;20(9):1054-60.
46. Minde DP, Anvarian Z, Rüdiger SG, Maurice MM. Messing up disorder: how do missense mutations in the tumor suppressor protein APC lead to cancer? *Mol Cancer.* 2011;10:101.
47. Crobach S, van Wezel T, Vasen HF, Morreau H. Ovarian metastases of colorectal and duodenal cancer in familial adenomatous polyposis. *Fam Cancer.* 2012;11(4):671-3.
48. Ruys AT, Alderlieste YA, Gouma DJ, Dekker E, Mathus-Vliegen EM. Jejunal cancer in patients with familial adenomatous polyposis. *Clin Gastroenterol Hepatol.* 2010;8(8):731-3.

49. Spirio L, Olschwang S, Groden J, Robertson M, Samowitz W, Joslyn G, et al. Alleles of the APC gene: an attenuated form of familial polyposis. *Cell*. 1993;75(5):951-7.
50. Nasioulas S, Jones IT, St John DJ, Scott RJ, Forrest SM, McKinlay Gardner RJ. Profuse familial adenomatous polyposis with an adenomatous polyposis coli exon 3 mutation. *Fam Cancer*. 2001;1(1):3-7.
51. Wanitsuwan W, Vijasika S, Jirattanasopa P, Horpaopan S. A distinct APC pathogenic germline variant identified in a southern Thai family with familial adenomatous polyposis. *BMC Med Genomics*. 2021;14(1):87.
52. Ercoskun P, Yuce Kahraman C, Ozkan G, Tatar A. Genetic Characterization of Hereditary Cancer Syndromes Based on Targeted Next-Generation Sequencing. *Mol Syndromol*. 2022;13(2):123-31.
53. Filipe B, Baltazar C, Albuquerque C, Fragoso S, Lage P, Vitoriano I, et al. APC or MUTYH mutations account for the majority of clinically well-characterized families with FAP and AFAP phenotype and patients with more than 30 adenomas. *Clin Genet*. 2009;76(3):242-55.
54. De Rosa M, Dourisboure RJ, Morelli G, Graziano A, Gutiérrez A, Thibodeau S, et al. First genotype characterization of Argentinean FAP patients: identification of 14 novel APC mutations. *Hum Mutat*. 2004;23(5):523-4.
55. Latchford A, Volikos E, Johnson V, Rogers P, Suraweera N, Tomlinson I, et al. APC mutations in FAP-associated desmoid tumours are non-random but not 'just right'. *Hum Mol Genet*. 2007;16(1):78-82.
56. Aretz S, Uhlhaas S, Sun Y, Pagenstecher C, Mangold E, Caspari R, et al. Familial adenomatous polyposis: aberrant splicing due to missense or silent mutations in the APC gene. *Hum Mutat*. 2004;24(5):370-80.
57. Friedl W, Aretz S. Familial adenomatous polyposis: experience from a study of 1164 unrelated german polyposis patients. *Hered Cancer Clin Pract*. 2005;3(3):95-114.
58. Li M, Gerber DA, Koruda M, O'Neil BH. Hepatocellular carcinoma associated with attenuated familial adenomatous polyposis: a case report and review of the literature. *Clin Colorectal Cancer*. 2012;11(1):77-81.
59. Yurgelun MB, Allen B, Kaldate RR, Bowles KR, Judkins T, Kaushik P, et al. Identification of a Variety of Mutations in Cancer Predisposition Genes in Patients With Suspected Lynch Syndrome. *Gastroenterology*. 2015;149(3):604-13.e20.
60. Ficari F, Cama A, Valanzano R, Curia MC, Palmirotta R, Aceto G, et al. APC gene mutations and colorectal adenomatosis in familial adenomatous polyposis. *Br J Cancer*. 2000;82(2):348-53.
61. Lagarde A, Rouleau E, Ferrari A, Noguchi T, Qiu J, Briaux A, et al. Germline APC mutation spectrum derived from 863 genomic variations identified through a 15-year medical genetics service to French patients with FAP. *J Med Genet*. 2010;47(10):721-2.
62. Gavert N, Yaron Y, Naiman T, Bercovich D, Rozen P, Shomrat R, et al. Molecular analysis of the APC gene in 71 Israeli families: 17 novel mutations. *Hum Mutat*. 2002;19(6):664.
63. Davidson S, Leshanski L, Rennert G, Eidelman S, Amikam D. Maternal mosaicism for a second mutational event--a novel deletion--in a familial adenomatous polyposis family harboring a new germ-line mutation in the alternatively spliced-exon 9 region of APC. *Hum Mutat*. 2002;19(1):83-4.
64. Tsukanov AS, Pospekhova NI, Shubin VP, Kuzminov AM, Kashnikov VN, Frolov SA, et al. Mutations in the APC gene in Russian patients with classic form of familial adenomatous polyposis. *Russian Journal of Genetics*. 2017;53(3):369-75.
65. Nieminen TT, Pavicic W, Porkka N, Kankainen M, Järvinen HJ, Lepistö A, et al. Pseudoexons provide a mechanism for allele-specific expression of APC in familial adenomatous polyposis. *Oncotarget*. 2016;7(43):70685-98.

TABLES, FIGURES & SUPPLEMENTARY MATERIAL

Figure 1 APC-specific criteria embedded in a reclassification algorithm

Figure 2 Reclassification workflow and suggested method of operation for ongoing VCEP activity

Figure 3(a) Revised classification of all APC variants in ClinVar and LOVD using the APC-specific criteria **(b)** Classification of all APC variants in the original database (O) and after their revised classification (R) based on the APC-specific criteria by variant type

Table 1 Further curation of selected variants remaining VUS after application of reclassification algorithm

Supplementary Table 1 Status of ClinVar and all APC variant databases

Supplementary Table 2

(a) Final classification of 191 structural variants in ClinVar (retrieved 21/03/22) and InSiGHT LOVD (retrieved 12/05/22) by APC specificications

(b) Final classification of 10038 non-structural variants in ClinVar (retrieved 21/03/22) and InSiGHT LOVD (retrieved 12/05/22) by APC specificications

(c) SpliceAI and MaxEntScan predictions for non-truncating APC variants

(d) APC variants with experimental data

(e) APC variants with clinical data

Supplementary Table 3 A list of variants reclassified as VUS from clinically relevant classifications (B/LB/P/LP)

Supplementary Table 4 Frequency of application of the APC-specific variant classification criteria

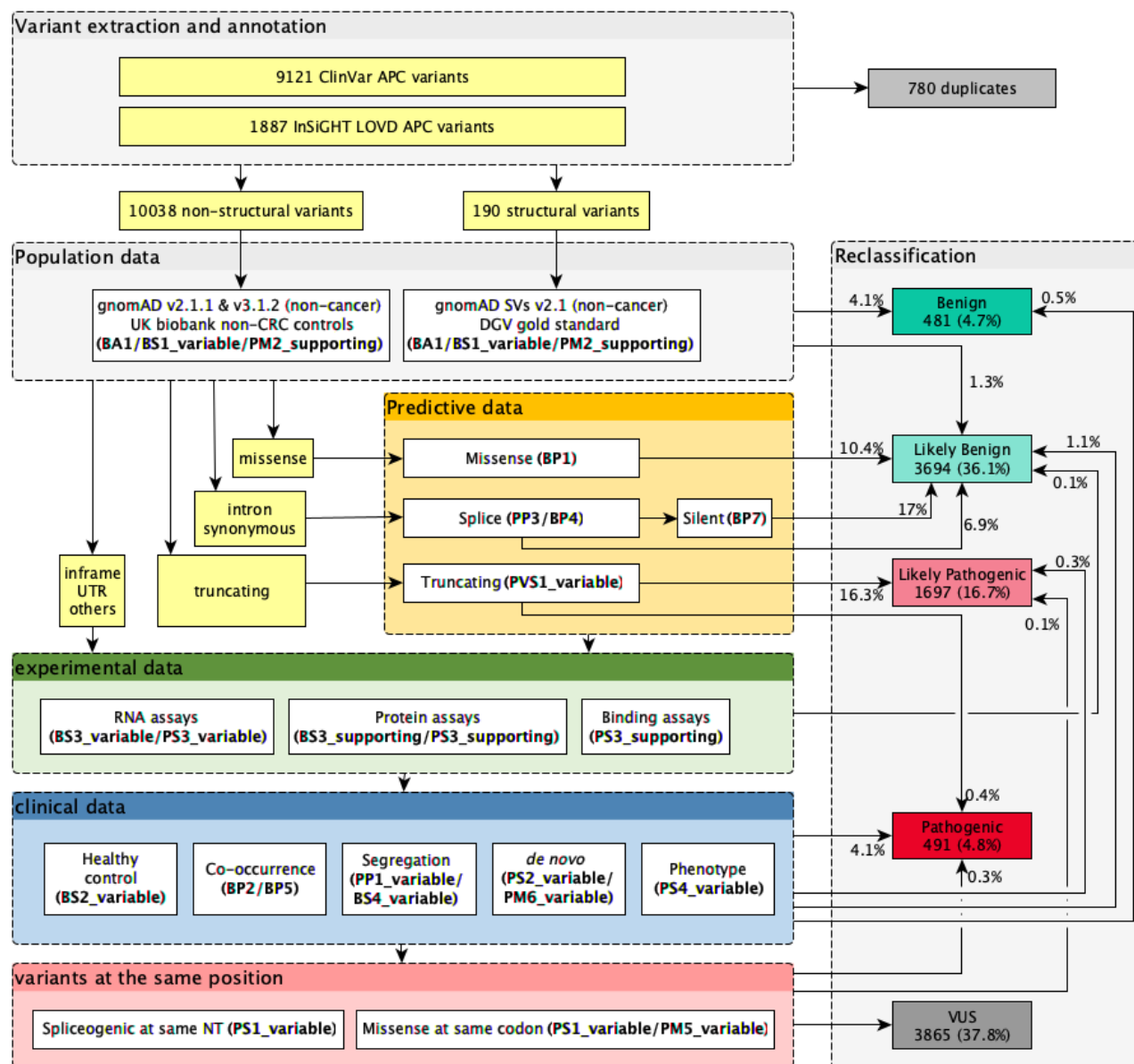


Figure 1 APC-specific criteria embedded in a reclassification algorithm An algorithm demonstrating the application of all eligible APC-specific codes to APC variants in ClinVar and the InSIGHT LOVD) in a stepwise approach, and the percentage of variants that reached a B/LB or P/LP classification at each step. Firstly, the highest minor allele frequency of non-structural and structural variants was calculated from gnomAD non-cancer datasets or UK Biobank non-colorectal cancer control data, and gnomAD SVs (structural variants) or DGV (database of genomic variants) gold standard, respectively. Predictive criteria were then applied based on the most severe variant consequence as predicted by Ensemble VEP (variant effect predictor). Variants with any experimental and/or clinical evidence were identified and assigned corresponding code. Finally, splice variants at the same nucleotide and missense variants at the same codon were identified and the variants at the same position criteria were applied.

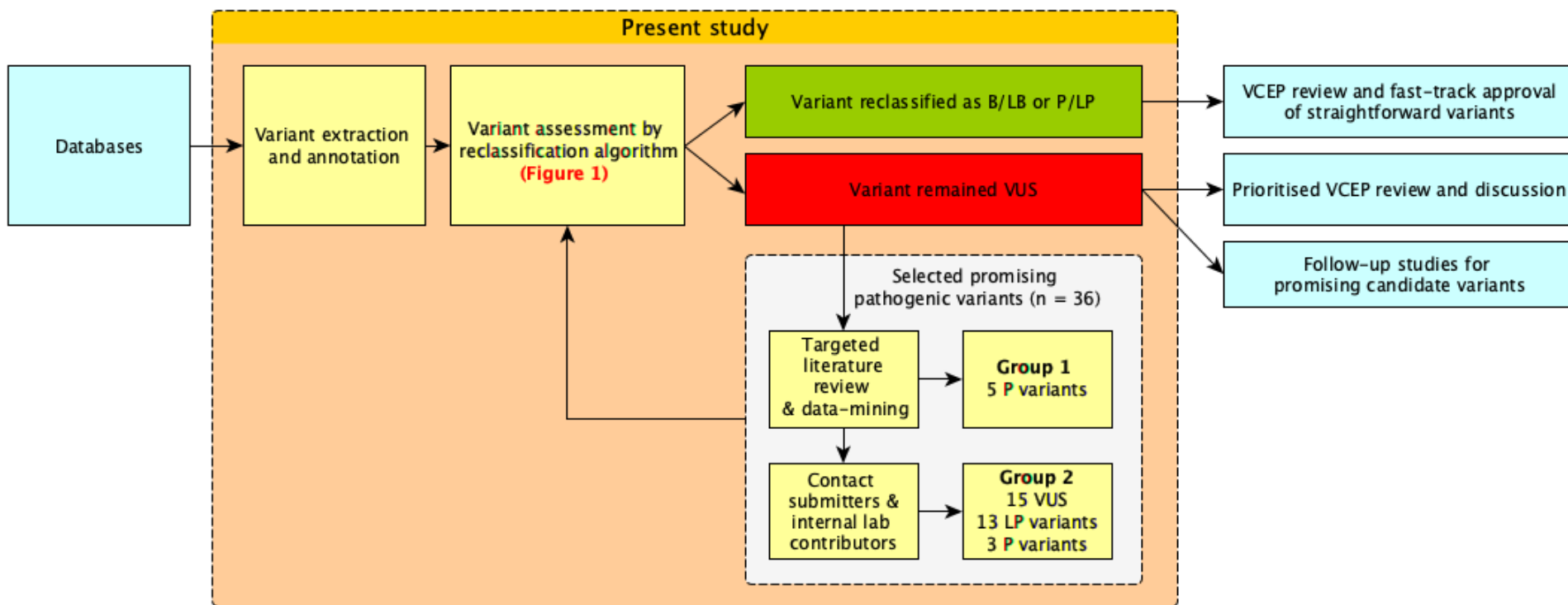
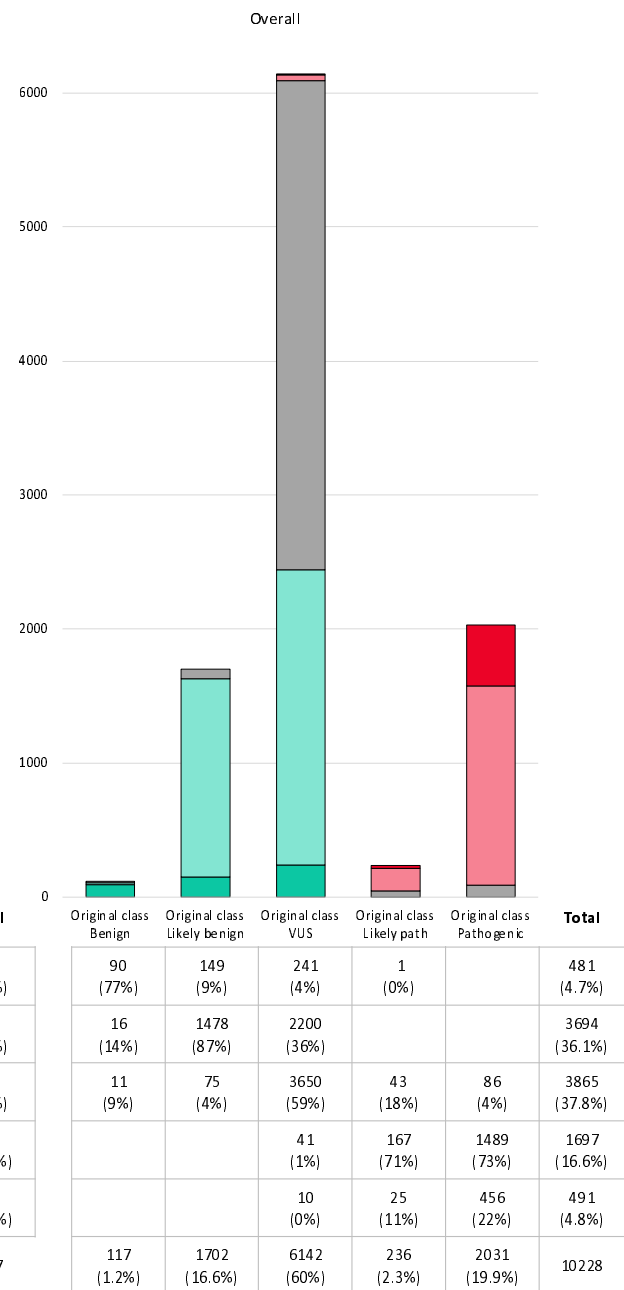
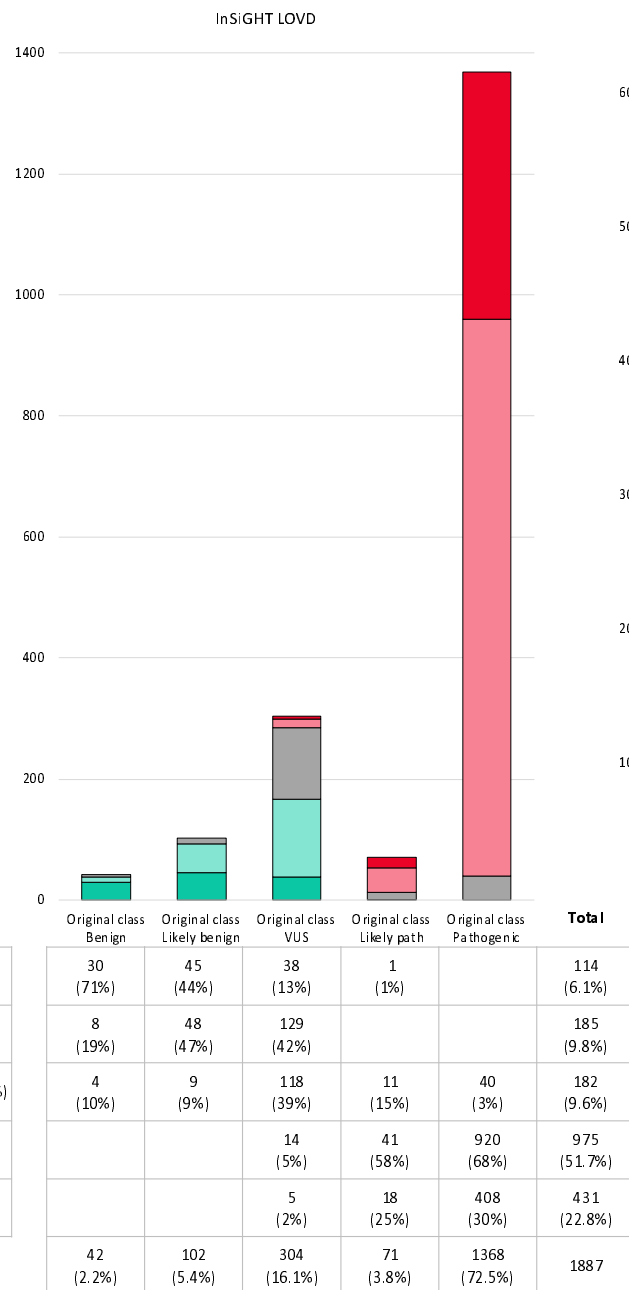
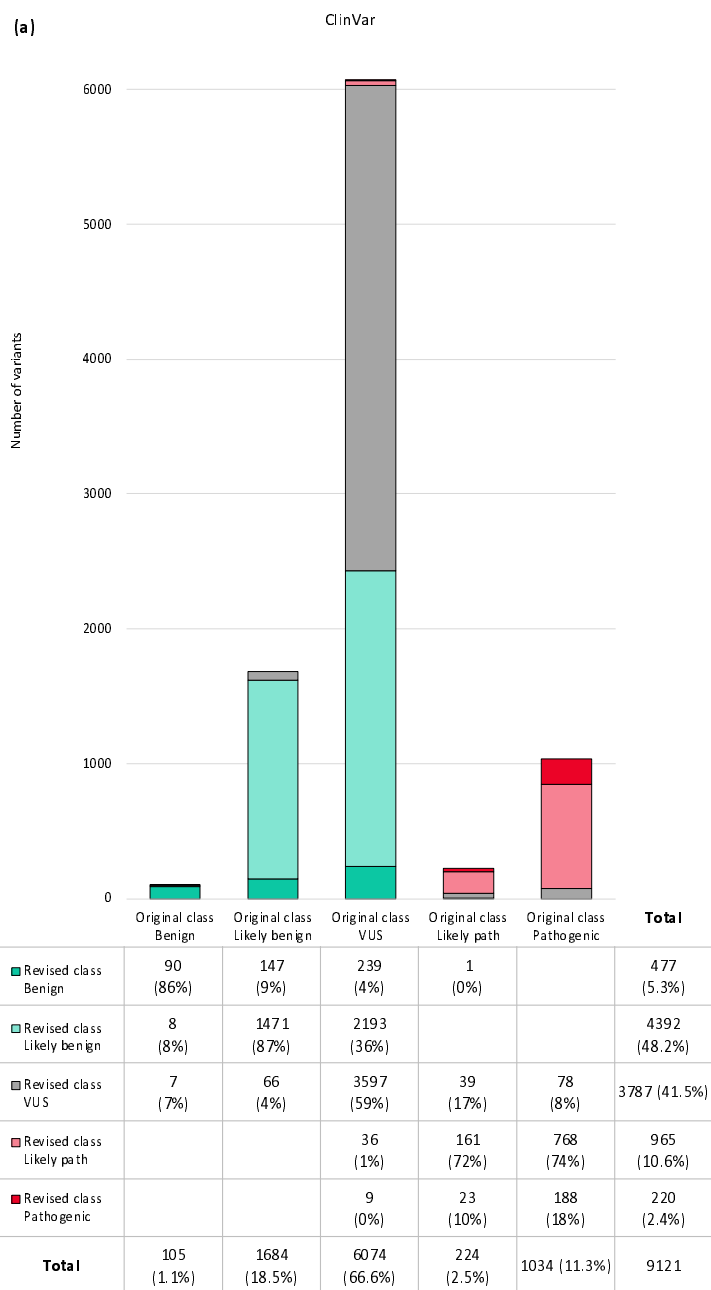


Figure 2. Reclassification workflow and suggested method of operation for ongoing VCEP activity This workflow summarised the organisational procedures undertaken in this study for variant reclassification, validation of the reassigned class and prioritisation of variants with high clinical importance for different modes of VCEP review and approval. Relatively straightforward variants might be processed in batch and become candidates for fast-track VCEP approval (e. g. variants fulfilling BA1 or BS1 plus BP1). From the remaining VUS 36 variants with some evidence for pathogenicity were selected and five variants were re-assessed as Pathogenic by a targeted literature review and data mining (Group 1). The remaining 31 variants were reassessed based on further clinical information requested from respective ClinVar submitters and internal lab contributors, which ultimately leads to prioritised VCEP review.

(a)



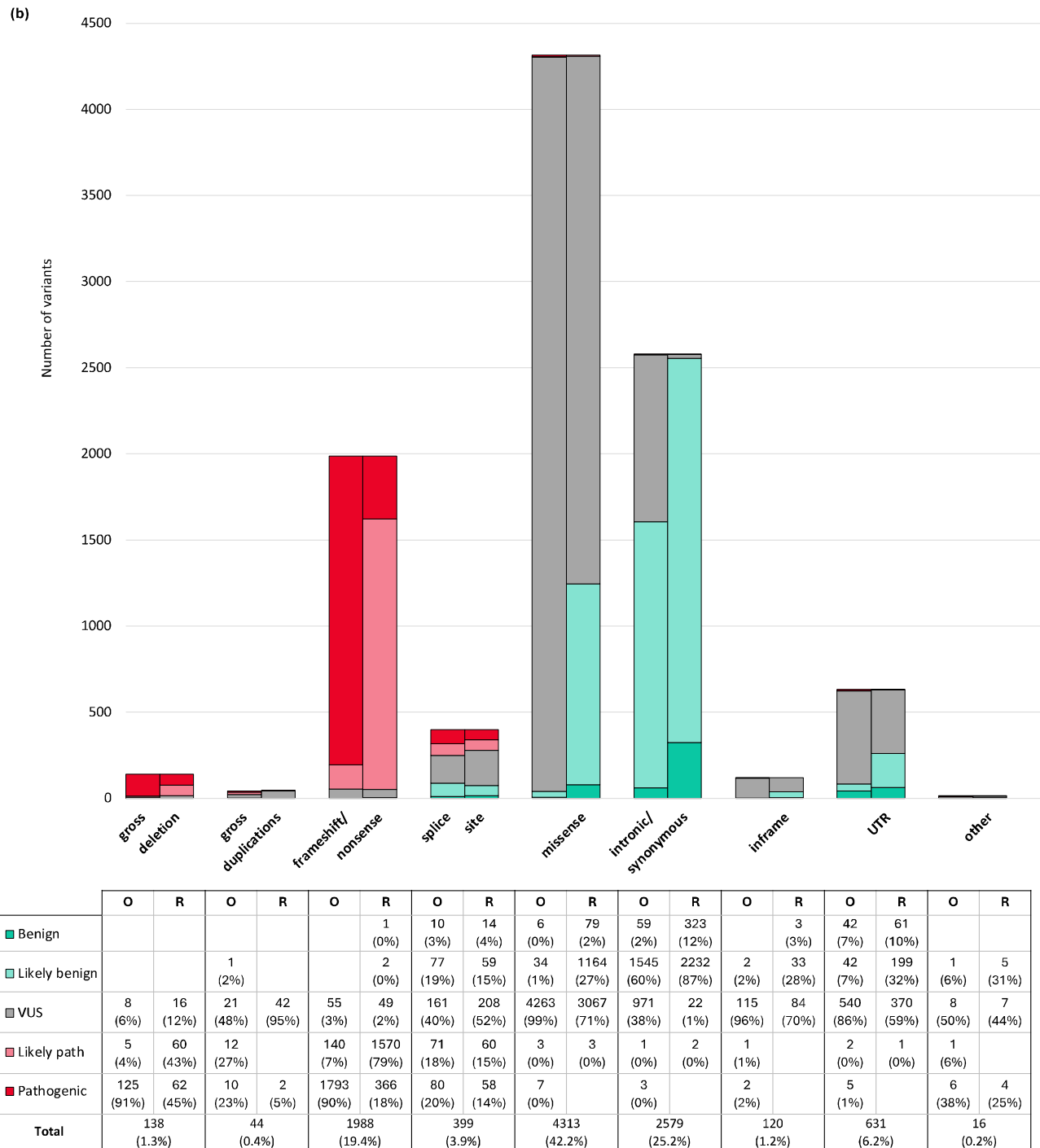


Figure 3 (a) Revised classification of all APC variants using the APC-specific criteria Each column shows the number of variants with their original assertions in ClinVar, InSIGHT LOVD and the overall dataset. The coloured segments of each column represent the revised classification using the APC-specific criteria. **(b) Classification of all APC variants in the original database (O) and their revised classification (R) by variant type** Variants are broadly categorised into nine categories: 138 gross deletions, 44 gross duplications, 1988 frameshift/nonsense, 399 splice site, 4313 missense, 2579 intronic/synonymous, 120 in-frame, 631 UTR, and 16 other variants, which included start-loss, stop-loss, stop-retained, Alu and SVA retrotransposon insertions, inversions and complex variant.

Table 1 Further curation of selected variants remaining VUS after application of reclassification algorithm

HGVSc	Database_ID	Prior classification	Criteria applied by algorithm	Classification by algorithm	Further curation	Final criteria applied	Final classification of VCEP
Group 1 Previously Pathogenic variants, for which further assessment based on literature review, data-mining and reassessment of minor allele frequency (MAF) criteria resulted in a classification of Pathogenic							
c.471G>A (p.Trp157Ter)	ClinVar 411479	Pathogenic	PVS1	VUS	1.5 phenotype points (47-50); revised MAF criteria	PVS1, PS4_supp, PM2_supp ^a	Pathogenic
c.1312+4_1312+19del	LOVD APC_001939	Pathogenic	BP4, PM2_supp	VUS	6.5 phenotype points, RNA assay result, segregated in 8 meiosis (51)	PS3_mod, PS4, PP1_strong, PM2_supp	Pathogenic
c.1333C>T (p.Gln445Ter)	ClinVar 438865	Pathogenic	PVS1, BS1	VUS	1 phenotype point (52, 53); revised MAF criteria	PVS1, PS4_supp, PM2_supp ^a	Pathogenic
c.2546_2551del (p.Asp849_Ser851delinsGly)	LOVD APC_000075	Pathogenic	PM2_supp, PS4_mod	VUS	Error in annotation: c.2546delATAGAAG (legacy name). Correct nomenclature: c.2546_2552del; p.(Asp849Valfs*10), 1 phenotype point, segregation in 3 meioses (11)	PVS1, PS4_supp, PM2_supp, PP1	Pathogenic
c.4669_4670del (p.Ile1557Ter)	ClinVar 183857	Pathogenic	PVS1	VUS	1 phenotype point (54, 55); revised MAF criteria	PVS1, PS4_supp, PM2_supp ^a	Pathogenic
Group 2 Promising potentially pathogenic VUS, further assessment based on clinical evidence from ClinVar submitters and internal lab contributors and reassessment of MAF criteria							
c.136-4A>G	925741	VUS	PP3, BS1	VUS	observed in 1 individual worth 0 phenotype points; revised MAF criteria	PP3	VUS
c.156del (p.Gly53GlufsTer17)	ClinVar 654864	Pathogenic	PVS1	VUS	observed in 1 individual worth 0 phenotype points; revised MAF criteria	PVS1, PM2_supp ^a	Likely pathogenic
c.203del (p.Leu68TyrfsTer2)	ClinVar 934724	Pathogenic	PVS1	VUS	observed in 4 individuals worth 1 phenotype point; revised MAF criteria	PVS1, PS4_supp, PM2_supp ^a	Pathogenic
c.220+2T>A	ClinVar 141515	Likely pathogenic	PVS1, BS1	VUS	observed in 35 individuals worth 8 phenotype points; revised MAF criteria	PVS1, PS4, PM2_supp ^a	Pathogenic
c.422G>C (p.Arg141Thr)	ClinVar 1056286	VUS	PVS1_strong, PM2_supp	VUS	observed in 3 individuals worth 0 phenotype points	PVS1_strong, PM2_supp	VUS
c.422G>A (p.Arg141Lys)	ClinVar 824696	VUS	PVS1_strong	VUS	observed in 0 individual worth 0 phenotype points; revised MAF criteria	PVS1_strong, PM2_supp	VUS
c.423-9A>G	ClinVar 469955	Pathogenic	PP3, PM2_supp, PS3_mod	VUS	observed in 4 individuals worth 2 phenotype points	PP3, PM2_supp, PS3_mod, PS4_mod	Likely pathogenic
c.531+5_531+8del	ClinVar 537529	Likely pathogenic	PP3, PM2_supp, PS3_mod, PS4_supp	VUS	observed in 5 individuals worth 3.5 phenotype points (56, 57)	PP3, PM2_supp, PS3_mod, PS4_mod	Likely pathogenic
c.531+5G>A	ClinVar 127305	Pathogenic	PP3, PM2_supp, PS4_supp, PS1_mod	VUS	observed in 3 individuals worth 1.5 phenotype points; 2.5 phenotype points in total (58, 59)	PP3, PM2_supp, PS4_mod, PS1_mod	Likely pathogenic
c.531+6T>C	ClinVar 576816	VUS	PP3, PM2_supp, PS3_mod	VUS	observed in 2 individuals worth 2 phenotype points	PP3, PM2_supp, PS3_mod, PS4_mod	Likely pathogenic
c.623A>G (p.Gln208Arg)	LOVD APC_000758	Pathogenic	BP1, PM2_supp, PP1	VUS	1 phenotype point (60)	BP1, PM2_supp, PS4_supp, PP1	VUS
c.645+2T>G	ClinVar 185659	Likely pathogenic	PVS1_mod, PM2_supp	VUS	observed in 2 individuals worth 1 phenotype point	PVS1_mod, PM2_supp, PS4_supp	VUS

c.835-17A>G	ClinVar 822326	Likely pathogenic	PM2_supp, PS3_mod	VUS	observed in 3 individuals worth 0 phenotype points	PM2_supp, PS3_mod	VUS
c.835-7T>G	ClinVar 433614	Likely pathogenic	PP3, PM2_supp, PS3_mod	VUS	observed in 1 individual worth 1 phenotype point	PP3, PM2_supp, PS3_mod, PS4_supp	VUS
c.933G>C (p.Lys311Asn)	1025291	Likely pathogenic	PVS1_supp, PM2_supp	VUS	observed in 2 individuals worth 1.5 phenotype points (61)	PVS1_strong, PM2_supp, PS4_supp	Likely pathogenic
c.1042C>T (p.?)	ClinVar 955439	Pathogenic	PVS1	VUS	1 phenotype point (62, 63); revised MAF criteria	PS4_supp, PM2_supp, BS2_supp	VUS
c.1312+3A>C	ClinVar 486792	Likely pathogenic	PP3, PM2_supp, PS1_mod	VUS	observed in 1 individual worth 1 phenotype point; 1.5 phenotype points in total (64)	PP3, PM2_supp, PS1_mod, PS4_supp	VUS
c.1312+5G>C	ClinVar 265372	Likely pathogenic	PM2_supp, PS4_supp, PS1_mod	VUS	observed in 2 individuals worth 2.5 phenotype points	PM2_supp, PS4_mod, PS1_mod	VUS
c.1408+735A>T	LOVD APC_001244	Pathogenic	PM2_supp, PS3_mod, PS4_supp	VUS	1 phenotype point (65)	PS4_mod, PS3_mod, PM2_supp	VUS
c.1409-5A>G	ClinVar 411406	Pathogenic	PP3, PM2_supp, PS3_mod, PS4_supp	VUS	observed in 5 individuals worth 3 phenotype points	PS3_mod, PS4_mod, PP3, PM2_supp	Likely pathogenic
c.1409-3T>G	ClinVar 485146	Likely pathogenic	PP3, PM2_supp, PS3_mod	VUS	observed in 4 individuals worth 2 phenotype points (61)	PS3_mod, PS4_mod, PP3, PM2_supp	Likely pathogenic
c.1743G>C (p.Lys581Asn)	ClinVar 428153	Likely pathogenic	PVS1_strong, PM2_supp	VUS	observed in 7 individuals worth 1 phenotype point	PVS1_strong, PS4_supp, PM2_supp	Likely pathogenic
c.1902T>G (p.Ser634Arg)	ClinVar 231954	Likely pathogenic	BP1, PS3_mod	VUS	observed in 2 individuals worth 0.5 phenotype points; revised MAF criteria and functional data	BP1, PM2_supp	VUS
c.3950A>G (p.Glu1317Gly)	ClinVar 1319598	VUS	BP1, PM2_supp	VUS	observed in 2 individuals worth 0 phenotype points	BP1, PM2_supp	VUS
c.4139C>T (p.Thr1380Ile)	ClinVar 233890	VUS	BP1, PM2_supp, PS4_supp	VUS	observed in 9 individuals worth 3 phenotype points	BP1, PM2_supp, PS4_mod	VUS
c.4735A>T (p.Ile1579Phe)	ClinVar 246402	VUS	BP1, PM2_supp, PS3_supp	VUS	observed in 4 individuals worth 1.5 phenotype points	BP1, PM2_supp, PS3_supp, PS4_supp	VUS
c.5038C>T (p.Gln1680Ter)	ClinVar 230520	Likely pathogenic	PVS1, BS1	VUS	observed in 1 individual worth 0 phenotype points; revised MAF criteria	PVS1, PM2_supp ^a	Likely pathogenic
c.6905C>G (p.Ser2302Ter)	ClinVar 428166	Pathogenic	PVS1	VUS	observed in 1 individual worth 0 phenotype points; revised MAF criteria	PVS1, PM2_supp ^a	Likely pathogenic
c.7489_7490insT (p.Ser2497PhefsTer14)	ClinVar 653103	Pathogenic	PVS1, BS1	VUS	observed in 4 individuals worth 1 phenotype point; revised MAF criteria	PVS1, PS4_supp, PM2_supp ^a	Pathogenic
c.7798_7801del (p.Gln2600ValfsTer15)	ClinVar 827255	Pathogenic	PVS1, BS1	VUS	observed in 2 individuals worth 0.5 phenotype points; observed in 3 individuals worth 3 healthy individual points; revised MAF criteria	PVS1, PM2_supp ^a , BS2_supp	Likely pathogenic
c.7803_7807del (p.Ser2601ArgfsTer17)	ClinVar 648862	Likely pathogenic	PVS1	VUS	observed in 1 individual worth 0 phenotype points; revised MAF criteria	PVS1, PM2_supp ^a	Likely pathogenic

^a Variants for which the reassessment of Minor allele frequency (MAF) criteria were relevant for the final evaluation as LP/P