Lack of evidence for CDK12 as an ovarian cancer predisposing gene

Alexandre Eeckhoutte, Mathilde Saint-Ghislain, Manon Reverdy, Virginie Raynal, Sylvain Baulande, Guillaume Bataillon, Lisa Golmard, Dominique Stoppa-Lyonnet, Tatiana Popova, Claude Houdayer, et al.

Lack of evidence for *CDK12* as an ovarian cancer predisposing gene

Alexandre Eeckhoutte[a,b], Mathilde Saint-Ghislain[a,b], Manon Reverdy[a,b], Virginie Raynal[b,c], Sylvain Baulande[c], Guillaume Bataillon[d], Lisa Golmard[e], Dominique Stoppa-Lyonnet[a,e,f], Tatiana Popova[a,b], Claude Houdayer[e§], Elodie Manié[a,b], Marc-Henri Stern[a,b,e]

[a]Inserm U830, DNA Repair and Uveal Melanoma (D.R.U.M.), Equipe labellisée par la Ligue Nationale Contre le Cancer, Institut Curie, 26 rue d'Ulm, 75248, Paris, France

[b]Inserm U830, Institut Curie, PSL Research University, 26 rue d'Ulm, 75248, Paris, France

[c]NGS platform, Institut Curie, PSL Research University, 26 rue d'Ulm, 75248, Paris, France

[d]Department of Biopathology, Institut Curie, PSL Research University, 26 rue d'Ulm, 75248, Paris, France

[e]Institut Curie, Hôpital, Service de Génétique, 26 rue d'Ulm, 75248, Paris, France

[f]University Paris Descartes, Sorbonne Paris Cité, 12 Rue de l'École de Médecine, 75006 Paris, France

§present address: Department of Genetics, and Normandy University, UNIROUEN, Inserm U1245, Normandy Centre for Genomic and Personalized Medicine, Rouen University Hospital, 37 Boulevard Gambetta, 76000 Rouen, France

Corresponding author: Marc-Henri Stern, Inserm U830, D.R.U.M., Institut Curie, 26 rue d'Ulm, 75248, Paris, France, Email: marc-henri.stern@curie.fr, Phone: +33 1 56 24 66 46

ORCID iD: 0000-0002-8100-2272

Keywords: CDK12, cancer susceptibility, ovarian carcinoma, pool sequencing, NGS

Abstract

CDK12 variants were investigated as a genetic susceptibility to ovarian cancer in a series of 416 unrelated and consecutive patients with ovarian carcinoma and who carry neither germline *BRCA1* nor *BRCA2* pathogenic variant. The presence of *CDK12* variants was searched in germline DNA by massive parallel sequencing on pooled DNAs. The lack of detection of deleterious variants and the observed proportion of missense variants in the series of ovarian carcinoma patients as compared with all human populations strongly suggests that *CDK12* is not an ovarian cancer predisposing gene.

Acknowledgments

Supported by the *Institut National de la Santé et de la Recherche Médicale* (INSERM) and the *Institut Curie*. M. S-G and A. E. are supported by fellowships from the *Université de Rouen* and the *Ligue Nationale Contre le Cancer*, respectively. The Institut Curie ICGex NGS platform is funded by the EQUIPEX *investissements d'avenir* program (ANR-10-EQPX-03) and ANR10-INBS-09-08 from the *Agence Nationale de la Recherch*e. We thank the patients. We also acknowledge support from the *Institut Curie* for sample collection, banking and processing: the Biological Resource Center and its members (O. Mariani).

Introduction

Epithelial ovarian carcinoma (EOC) is a rare but dreadful disease and represents the 5th cause of death by cancer in women worldwide. The most frequent histology of this carcinoma is High Grade Serous Ovarian Carcinoma (HGSOC). The known risks factors of EOC are age, hormonal history and genetic factors [1]. Genetic factors are involved in approximately 15% of HGSOC, often associated with the so-called hereditary breast and ovarian cancer syndrome (HBOC). *BRCA1* and *BRCA2*, which code key actors of the homologous recombination (HR) DNA repair pathway, are the two major cancer predisposing genes. Heterozygote germline mutations of *BRCA1* and *BRCA2* lead to the most important increased risk to develop an HGSOC with a Relative Risk (RR) of 40 and 18, respectively [2]. They explain 65% to 75% of hereditary EOC. In addition, mutations of other genes belonging to the HR pathway, such as *RAD51* paralogs, also participate to the HBOC syndrome [3-5]. A second group of predisposition genes belong to the MMR family (*MLH1*, *MSH2*, *MSH6*, *PMS2*), mutated in the Lynch syndrome. Lynch syndrome is associated with a RR of [3.6-13] for EOC, mainly of the endometrioid and clear cell subtypes, and explains ~10% of hereditary EOC [1, 6-8]. Strikingly, all these predisposition genes encode proteins involved in DNA maintenance. Recently, *CDK12* emerged as an important player in ovarian carcinoma. *CDK12* (cyclin-dependent kinase 12) is one of the ten most frequently mutated genes in HGSOC (3% of the TCGA cohort) and is also mutated in 7% of metastatic castration-resistant prostate cancer [9]. *CDK12* behaves as a classical tumor suppressor gene with bi-allelic somatic inactivation in tumors, with in most cases one deleterious mutation on one allele and one chromosomal partial deletion evidenced by loss of heterozygosity (LOH). We have previously shown that *CDK12*-inactivated tumors are associated with an unusual form of genomic instability named the TD-plus phenotype and characterized by hundreds of large tandem duplications of up to 10 megabases in size [10]. *CDK12* is an essential gene during development as *Cdk12* inactivation is embryonic lethal in mouse models [11]. CDK12 is a nuclear serine threonine kinase that dimerizes with Cyclin K (CCNK). Until recently, the only known targets of

CCNK/CDK12 were serines 2 and 5 of the carboxy-terminal-domain (CTD) of the RNA polymerase II, required for elongation and end termination of transcription. In *in-vitro* studies, CDK12 is required for the expression of a subset of DNA Damage Response (DDR) genes, including *BRCA1, FANCI, FANCD2* [12], and conversely, *CDK12*-inactivated cell models are highly sensitive to PARP inhibitors [13]. Both *in vitro* studies and analyses of *CDK12*-mutated tumors strongly suggest that CDK12 plays a role in genomic maintenance. Recent studies have showed that CDK12 acts by suppressing intronic polyadenylation events, including in DNA repair genes [14-16]. CDK12 phosphorylates 4E-BP1 to enable mTORC1-dependent translation and maintains mitotic genome stability [17].

CDK12 as an important tumor suppressor gene in ovarian tumorigenesis pointed it as a potential predisposition gene for EOC. We explored this hypothesis by investigating the germline status of *CDK12* in a series of 416 unselected consecutive and unrelated patients with EOC, negative for *BRCA1* or *BRCA2* mutations.

Material and methods

Patients

A series of blood DNA from 416 unselected consecutive and unrelated patients was assembled from the Genetic Department of Institut Curie, initially explored negative for *BRCA1* or *BRCA2* deleterious germline mutations, using the current techniques at time of diagnosis or reanalysis. Five of them were subsequently found to carry deleterious mutations of *BRCA2* (1 case), *RAD51C* (1), *RAD51D* (1), *PMS2* (1) and *TP53* (1). All patients had personal history of EOC (mean age at diagnosis: 56 years-old) and benefited from genetic counseling. One hundred twenty-three of these patients had also developed one or more breast cancers. All patients have signed an informed consent for research of new cancer predisposing genes.

CDK12 sequencing in pooled DNA and positive control pools

Germline DNAs of the 416 patients were quantified using Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific). The 416 DNAs were pooled into 52 equimolar pools of eight DNAs per pool, with an expected variant frequency of one mutated allele in 16 alleles (6.25%). Two additional pools were constituted as positive controls, containing seven patients' DNAs plus one tumor DNA with a known *CDK12* mutation, c.137del/p.Lys46SerfsX11 and c.212dup/p.Glu72GlyfsX3 for pools #53 and #54, respectively. *CDK12* coding sequence and flanking introns were sequenced using the TruSeq Custom Amplicon Low Input Kit (Illumina). Briefly, the design included 32 amplicons of 250 bp for a theoretical coverage of 100% on a cumulative target of 4.61 kb. The library was produced by PCR and ligation from 20ng of pooled genomic DNA, barcoded with 54 indexes, quantified (Bioanalyzer, Agilent) and pooled in an equimolar ratio. The library was then paired-end sequenced (PE250) with a MiSeq v2 Nano flow cell (Illumina).

Bioinformatics pipeline

Quality control was performed using FASTQC. Reads were aligned to the hg19 assembly with BWA MEM (v. 0.7.5a). Primers were soft-clipped with BAMclipper [18]. BAM files were pre-processed with indels realignment and base quality score recalibration according to the GATK Best Practices (v. 3.5) [19]. Variants were detected by Samtools mpileup (v. 1.7) [20], HaplotypeCaller (v. 3.5) and Mutect2 (v. 3.5) . The union of all the variants called was annotated with ANNOVAR [21] according to different databases: EnsGene, COSMIC88, dbSNP151 and maximum allele frequency from 1000G, ExAC, ESP6500 and CG46. Variants were filtered out if: (i) synonymous, (ii) intronic and UTR located, (iii) biased for strand direction (outside of [0.4-0.6] ratio in a balanced site), (iv) frequency higher than 1% in any human reference population, (v) frequency higher than 1% in the patients' series, (vi) Variant Allele Frequency (VAF) < 1% and (vii) existing at a homozygous state in any individual. The remaining variants were then manually checked on IGV and five of them were discarded. A pileup approach was also implemented, with Samtools (v. 1.8) and a customized script to

retain indels supported by more than five reads (Figure 1). An *in silico* prediction of splice defects was also performed using the MaxEntScan tool (MaxEnt; [22]).

Variant validation

The eight DNAs from each positive pool were analyzed independently by Sanger sequencing for the identified variant. Briefly, PCR was performed from 50 ng of DNA using specific primers and Taq Gold using standard protocols (Primers and conditions available on request) and sequenced using Big Dye Terminator kit V1 (3130XL, Applied Biosystems). Quality control of the electropherograms was performed using FinchTV (PerkinElmer) and sequences were analyzed using SeqScape (Applied Biosystems).

Prediction

In-silico predictions of deleterious consequences of the non-synonymous variants were performed using CADD (Combined Annotation Dependent Depletion) Phred score, SIFT (Sorting Intolerant from Tolerant), Polyphen-2 and VEP (Variant Effect Predictor).

CDK12 defect genomic signature

Tumor DNA extracted from frozen or FFPE blocs from *CDK12*-variant carriers were obtained from the institutional Biobank and the Department of Pathology, respectively. Genomic profiling was obtained by shallow Whole Genome Sequencing (sWGS, approximatively 1 read per base) of the tumor DNA on NovaSeq (Illumina). Adapters were trimmed with Cutadapt (v. 1.18). The number of reads in windows of 10 kb was extracted and normalized for GC content and mappability with ControlFREEC [23]. The CDK12 TD plus pattern characteristic of *CDK12* inactivation was visually checked [10].

Statistical power of the study

Assuming a distribution in *CDK12* variants following a Poisson law, the theoretical frequency to detect at least one deleterious variant in a cohort of 927 cases with a power of 80 was calculated as the following:

$P(X \geq 1, \lambda) = 0.80 = 1 - P(X=0, \lambda)$

$$\lambda = \frac{-ln(0.20)}{927} = 1.7e-3$$

Results

The goal of this study was to evaluate *CDK12* as a potential EOC predisposing gene. We thus defined the frequency of *CDK12* germline variants in a series of 416 consecutive patients with ovarian carcinomas. The Region Of Interest (ROI) included 4525 bp corresponding to the coding sequence of *CDK12,* including 14 exons plus 2 base pairs of splicing sites. A pooled DNA sequencing approach was performed. The mean read depth on ROI was 920x and the majority of ROI (92%) displayed more than 320X coverage, corresponding to 20X per allele. The lowest depth on ROI was ranging from 83X to 227X in the different pools (mean 143x). The less covered regions were parts of exons 1 and 2, and the whole exon 10.

In addition to the two positive controls, a list of 151 variants was called by at least one of the variant calling methods. 21 different variants in 17 different pools were retained for validation after filtering. Ten predicted variants were not confirmed by Sanger sequencing, and thus were considered as false calls, whereas eleven of these variants were validated in 11 different patients, among which 10 were single nucleotide variants (SNVs) and one an in-frame 3-base deletion (Figure 1, Table 1). Out of these 11 variants, 9 were previously reported in the dbSNP database v151. All SNVs were missense variants with a predictive deleteriousness ranging from benign to moderate. One yet unreported SNP, c.A2712T, changes a glutamic acid codon conserved in all sequenced vertebrates up to lamprey and located within the kinase domain. However, the consequence of this change for aspartic acid was considered as mild.

No Loss of Function (LoF) variant, such as premature stop-gain, frameshift or splicing variant, was found in this series.

As compared with reported frequencies of these variants in the representative non-Finnish European population in the GnomAD database, none of these variants were significantly enriched in our series of EOC patients (Table 1). As *CDK12* deleterious variants are embryonic-lethal at homozygous state, we considered only the missense SNPs reported in dbSNP151 and never found at homozygous state in any human population. The proportion of such missense SNPs was not significantly different in our series from that of the representative non-Finnish European population (10/832 alleles and 823/113650 alleles respectively; Fisher's exact test : $p = 0.1453$), which is in accordance with *CDK12* not being an ovarian cancer predisposing gene.

Although no strong evidence supported the pathogenic effect of the identified variants, we further explored whenever possible the tumor of the corresponding variant carrier. We retrieved four EOC cases for whose tumor material was available. Given that *CDK12* is a tumor suppressor gene, we first assessed the loss of the wild-type allele in tumors, according to the Knudson/two-hit hypothesis. Only one of the four EOC had a loss of the wild allele, and the three others retained the wild-type allele in the tumors. A key feature of *CDK12*-inactivated tumors is a striking genomic profile enriched in numerous and very large tandem duplications, the TD-plus genomic signature [10], which was not found in the genomic profiles of the four tested EOCs with *CDK12* variants (Table 1 and Figure 2).

Altogether, we found no evidence of deleterious *CDK12* germline variants in our series of ovarian carcinoma, and so no evidence for its role as an ovarian susceptibility gene in the studied population.

Discussion

CDK12 recently emerged to play an important role in ovarian and prostatic carcinomas, as a tumor suppressor gene contributing to malignant transformation and genomic instability when inactivated. As such, *CDK12* was a good candidate to also play a role in cancer predisposition. This study evaluated the incidence of *CDK12* germline variants in a series of 416 unrelated and consecutive patients with ovarian carcinoma. A total of eleven *CDK12* exonic variants were identified by massive parallel sequencing and validated by Sanger sequencing. None of the variants was a Loss of Function variant (LoF). However, one was in the kinase domain on a well conserved codon up to lamprey, but with a mild acid to acid change of coded amino-acid. Unfortunately, no tumor sample was available for further investigation of this case. We then compared the proportion of missense variants found in our series from that of the representative non-Finnish European population and found no statistically significant difference. We further mined four variants, for which tumor samples were retrieved and no evidence of *CDK12* inactivation was found in these tumors. Altogether, we found no evidence for a role of *CDK12* as an ovarian cancer predisposition gene in an unbiased series of 416 *BRCA1/2*-wild-type patients with ovarian carcinoma. Furthermore, the analysis of The Cancer Genome Atlas (TCGA) series of 511 ovarian serous cystadenocarcinomas identified 15 cases with the TD-plus phenotype and bi-allelic inactivation of *CDK12*, none of which carrying any deleterious germline variant [10 and Popova *et al.*, unpublished data]. These analyses of combined series (TCGA and in-house cases) had a power of 80% to detect at least one deleterious germline mutation in the hypothesis of a deleterious variant frequency of $1.7\ 10^{-3}$ in EOC patients.

Interestingly, the number of LoF variants in Gnomad (n=6) is largely below the expected one, with a probability of being loss-of-function intolerant (pLI) score at 1 [24], and an observed / expected (oe) score of 0.05 (gnomad.broadinstitute.org). This suggests that LoF variants are counter-selected in human populations, even at the heterozygous state. Interestingly, a report described the existence of a deleterious c.1047-2A>G germline *CDK12* variant in 8 of the 106

HBOC cases tested (7.6%) in the Tatar population [25], but the association with HBOC was not confirmed in a replication study [26]. If replicated, this would suggest that *CDK12* variants could play a role in HBOC predisposition in some human populations, although strongly counter-selected in most human populations. The reason of this counter-selection is not clear. *Cdk12* inactivation is embryonic lethal in mouse models, but heterozygous *Cdk12*$^{\Delta/wt}$ pups are viable and born with the expected frequency [11]. Clearly, long-term follow-up of these *Cdk12*$^{\Delta/wt}$ mice may be instrumental to unravel the mechanism of intolerance of LoF variants in Humans. In a more distant model in Drosophila, a decline of courtship learning was observed in *CDK12* heterozygous flies [27]. This could be a plausible mechanism to explain the counter-selection of human heterozygous *CDK12*-mutant carriers, but caution should be taken given the evolutionary distance between flies and mammals.

In conclusion, our data evidenced the absence of deleterious *CDK12* variants in patients with ovarian carcinoma, confirming the rarity of such variants in the general population, and making unlikely the existence of deleterious variants in more than 0.2% of EOC patients. Thus our data do not support the role of *CDK12* in ovarian carcinoma susceptibility. The origin of the intolerance of deleterious *CDK12* variants in the population has yet to be explained.

Conflicts of interest

E. Manié, T. Popova and M.-H. Stern are named inventors of a patent licensed to Myriad Genetics. No potential conflicts of interest were disclosed by the other authors.

Author contributions

Conception and design: E. Manié, C. Houdayer, M.-H. Stern

Development of methodology: A. Eeckhoutte

Acquisition of data: E. Manié, A. Eeckhoutte, M. Saint-Ghislain, M. Reverdy, V. Raynal, S. Baulande, T. Popova, G. Bataillon, L. Golmard

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): E. Manié, A. Eeckhoutte, M.-H. Stern

Writing, review, and/or revision of the manuscript: E. Manié, D. Stoppa-Lyonnet, A. Eeckhoutte, M.-H. Stern

Study supervision: D. Stoppa-Lyonnet, S. Baulande, C. Houdayer, M.-H. Stern

References

1. Lheureux S, Gourley C, Vergote I, Oza AM (2019) Epithelial ovarian cancer. Lancet 393:1240-53 https://doi.org/10.1016/S0140-6736(18)32552-2

2. Kuchenbaecker KB, Hopper JL, Barnes DR, et al. (2017) Risks of Breast, Ovarian, and Contralateral Breast Cancer for BRCA1 and BRCA2 Mutation Carriers. JAMA 317:2402-16 https://doi.org/10.1001/jama.2017.7112

3. Golmard L, Castera L, Krieger S, et al. (2017) Contribution of germline deleterious variants in the RAD51 paralogs to breast and ovarian cancers. Eur J Hum Genet 25:1345-53 https://doi.org/10.1038/s41431-017-0021-2

4. Loveday C, Turnbull C, Ramsay E, et al. (2011) Germline mutations in RAD51D confer susceptibility to ovarian cancer. Nat Genet 43:879-82 https://doi.org/10.1038/ng.893

5. Loveday C, Turnbull C, Ruark E, et al. (2012) Germline RAD51C mutations confer susceptibility to ovarian cancer. Nat Genet 44:475-6; author reply 6 https://doi.org/10.1038/ng.2224

6. Bewtra C, Watson P, Conway T, Read-Hippee C, Lynch HT (1992) Hereditary ovarian cancer: a clinicopathological study. Int J Gynecol Pathol 11:180-7

7. Helder-Woolderink JM, Blok EA, Vasen HF, Hollema H, Mourits MJ, De Bock GH (2016) Ovarian cancer in Lynch syndrome; a systematic review. Eur J Cancer 55:65-73 https://doi.org/10.1016/j.ejca.2015.12.005

8. Bonadona V, Bonaiti B, Olschwang S, et al. (2011) Cancer risks associated with germline mutations in MLH1, MSH2, and MSH6 genes in Lynch syndrome. JAMA 305:2304-10 https://doi.org/10.1001/jama.2011.743

9. Wu YM, Cieslik M, Lonigro RJ, et al. (2018) Inactivation of CDK12 Delineates a Distinct Immunogenic Class of Advanced Prostate Cancer. Cell 173:1770-82 e14 https://doi.org/10.1016/j.cell.2018.04.034

10. Popova T, Manie E, Boeva V, et al. (2016) Ovarian Cancers Harboring Inactivating Mutations in CDK12 Display a Distinct Genomic Instability Pattern Characterized by Large

Tandem Duplications. Cancer Res 76:1882-91 https://doi.org/10.1158/0008-5472.CAN-15-2128

11. Juan HC, Lin Y, Chen HR, Fann MJ (2016) Cdk12 is essential for embryonic development and the maintenance of genomic stability. Cell Death Differ 23:1038-48 https://doi.org/10.1038/cdd.2015.157

12. Blazek D, Kohoutek J, Bartholomeeusen K, et al. (2011) The Cyclin K/Cdk12 complex maintains genomic stability via regulation of expression of DNA damage response genes. Genes Dev 25:2158-72 https://doi.org/10.1101/gad.16962311

13. Bajrami I, Frankum JR, Konde A, et al. (2014) Genome-wide profiling of genetic synthetic lethality identifies CDK12 as a novel determinant of PARP1/2 inhibitor sensitivity. Cancer Res 74:287-97 https://doi.org/10.1158/0008-5472.CAN-13-2541

14. Tien JF, Mazloomian A, Cheng SG, et al. (2017) CDK12 regulates alternative last exon mRNA splicing and promotes breast cancer cell invasion. Nucleic Acids Res 45:6698-716 https://doi.org/10.1093/nar/gkx187

15. Krajewska M, Dries R, Grassetti AV, et al. (2019) CDK12 loss in cancer cells affects DNA damage response genes through premature cleavage and polyadenylation. Nat Commun 10:1757 https://doi.org/10.1038/s41467-019-09703-y

16. Dubbury SJ, Boutz PL, Sharp PA (2018) CDK12 regulates DNA repair genes by suppressing intronic polyadenylation. Nature 564:141-5 https://doi.org/10.1038/s41586-018-0758-y

17. Choi SH, Martinez TF, Kim S, et al. (2019) CDK12 phosphorylates 4E-BP1 to enable mTORC1-dependent translation and mitotic genome stability. Genes Dev 33:418-35 https://doi.org/10.1101/gad.322339.118

18. Au CH, Ho DN, Kwong A, Chan TL, Ma ESK (2017) BAMClipper: removing primers from alignments to minimize false-negative mutations in amplicon next-generation sequencing. Sci Rep 7:1567 https://doi.org/10.1038/s41598-017-01703-6

19. DePristo MA, Banks E, Poplin R, et al. (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 43:491-8 https://doi.org/10.1038/ng.806

20. Li H, Handsaker B, Wysoker A, et al. (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078-9 https://doi.org/10.1093/bioinformatics/btp352

21. Wang K, Li M, Hakonarson H (2010) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 38:e164

22. Yeo G, Burge CB (2004) Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. J Comput Biol 11:377-94 https://doi.org/10.1089/1066527041410418

23. Boeva V, Popova T, Bleakley K, et al. (2012) Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data. Bioinformatics 28:423-5 https://doi.org/10.1093/bioinformatics/btr670

24. Lek M, Karczewski KJ, Minikel EV, et al. (2016) Analysis of protein-coding genetic variation in 60,706 humans. Nature 536:285-91 https://doi.org/10.1038/nature19057

25. Brovkina OI, Shigapova L, Chudakova DA, et al. (2018) The Ethnic-Specific Spectrum of Germline Nucleotide Variants in DNA Damage Response and Repair Genes in Hereditary Breast and Ovarian Cancer Patients of Tatar Descent. Front Oncol 8:421 https://doi.org/10.3389/fonc.2018.00421

26. Bogdanova NV, Schurmann P, Valova Y, et al. (2019) A Splice Site Variant of CDK12 and Breast Cancer in Three Eurasian Populations. Front Oncol 9:493 https://doi.org/10.3389/fonc.2019.00493

27. Pan L, Xie W, Li KL, et al. (2015) Heterochromatin remodeling by CDK12 contributes to learning in Drosophila. Proc Natl Acad Sci U S A 112:13988-93 https://doi.org/10.1073/pnas.1502943112

Figure legends

Figure 1. Variant calling workflow. Workflow of the *CDK12* pool targeted-sequencing, variant calling, filtering and validation. VCF: variant calling format; UTR: untranslated region; VAF: variant allele frequency in the pool; MAF: mutation allele frequency in population; IGV: Interactive Genome Viewer.

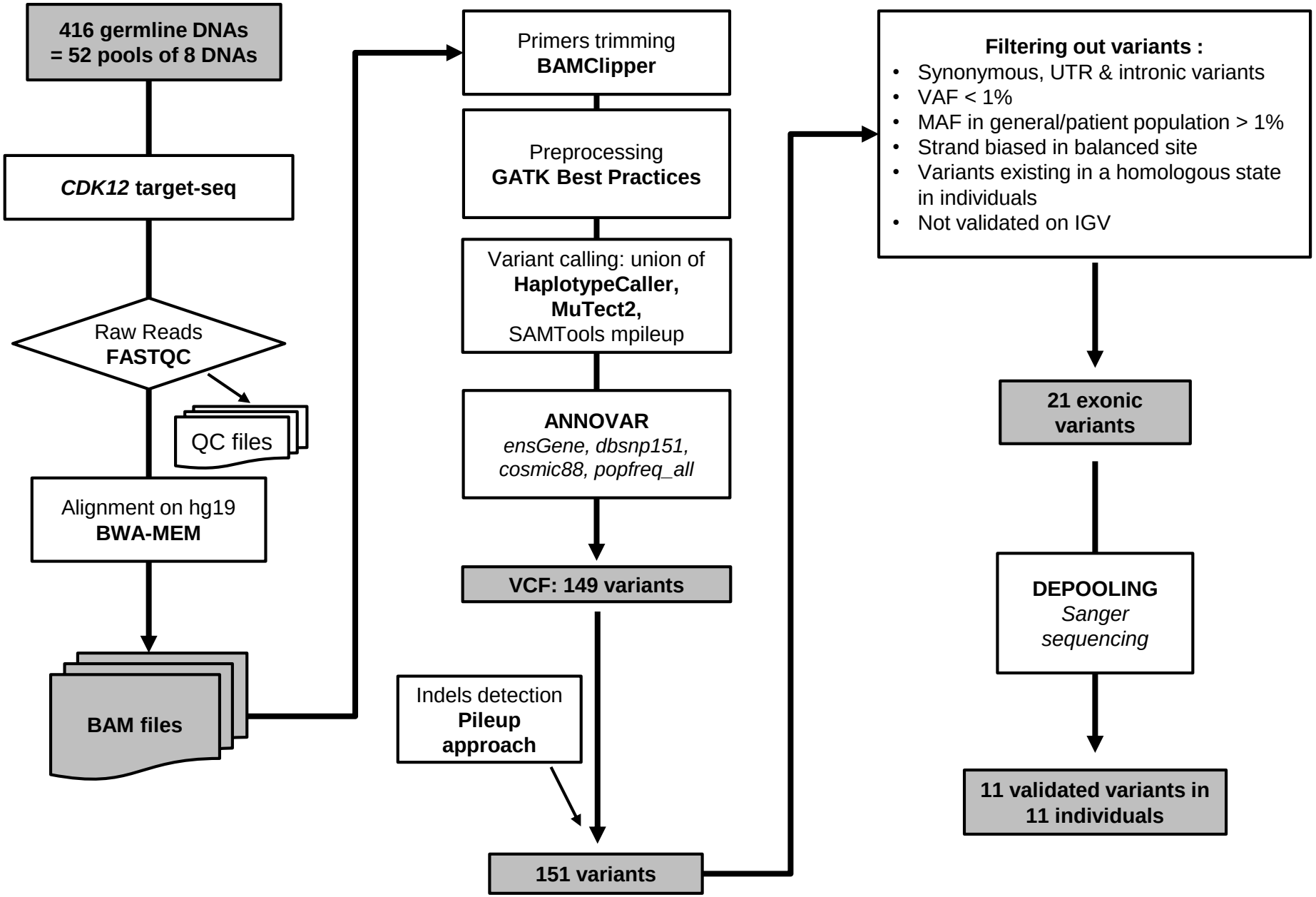
Figure 2. Genome profiling of epithelial ovarian carcinoma carrying *CDK12* variants. Genome profiling generated using low coverage whole genome sequencing (shallow WGS). Left panel: four epithelial carcinomas in patients carrying germline *CDK12* variants. Right panel: From top to bottom, examples of tumor genome profiles with Homologous Recombination Proficient (HRP), Homologous Recombination Deficient (HRD) and CDK12-inactivated TD-plus profiles, respectively.

Table legends

Table 1. Description of *CDK12* variants. Ref: reference allele; Alt: alternative allele; Pop max freq: maximum allele frequency in any human population (Non-Finish European – NFE - by default); AFR: AFRican population; EAS: East-Asian; EOC: Epithelial Ovarian Carcinoma; LOH: Loss Of Heterozygosity; ROH: Retention Of Heterozygosity; NA: Not Available; sWGS: shallow Whole Genome Sequencing; TD-Plus: tandem duplication-plus genome profile, characteristic of *CDK12* inactivation.

Figure 1

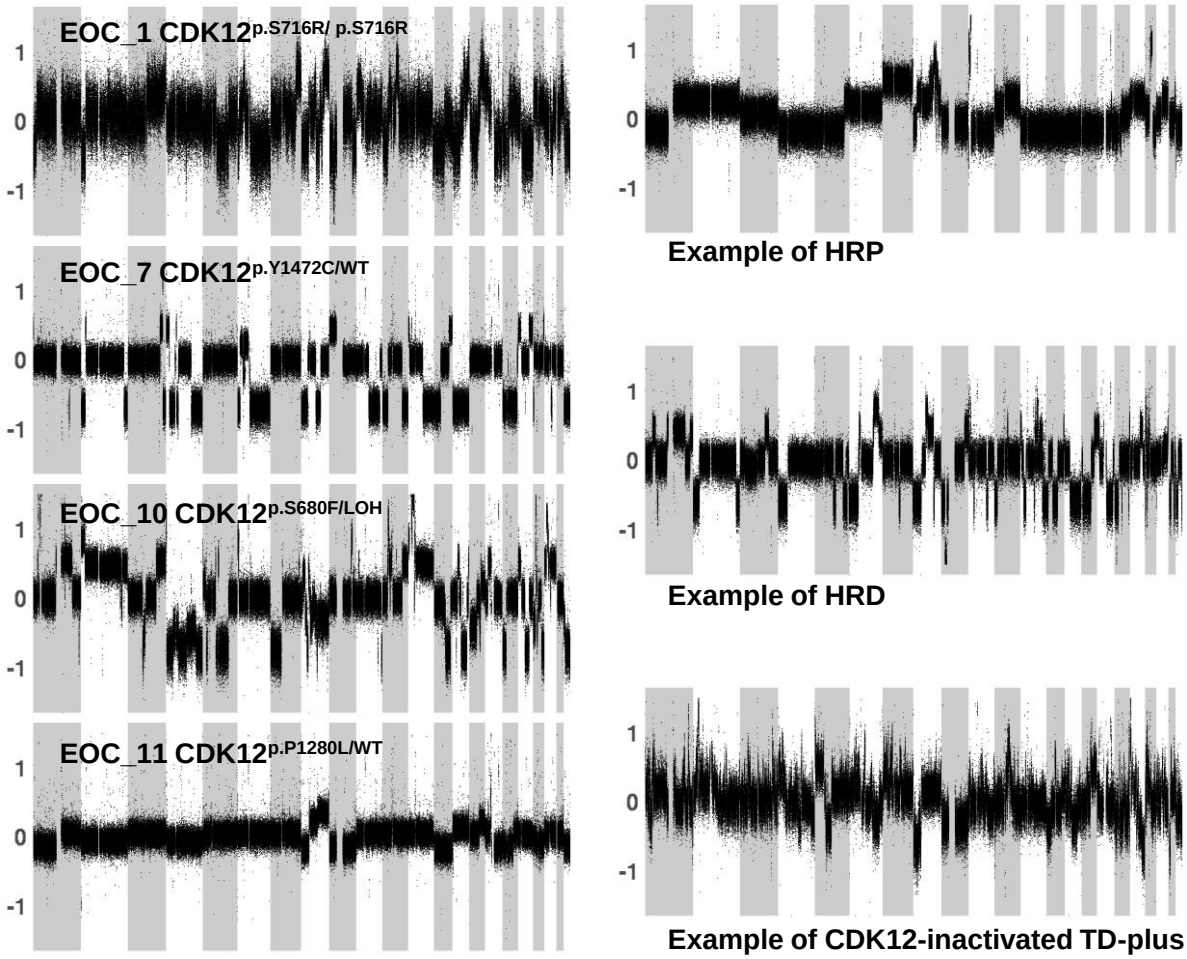
Click here to access/download;Figure;Fig1_workflow_complet_ae.pptx



Eeckhoutte et al. Figure 1

Figure 2

Click here to access/download;Figure;Fig2_sWGS_CDK12_article.ppt



Eeckhoutte et al. Figure 2

Table 1. Description of *CDK12* variants.

Samples	Coordinates	Ref	Alt	Variants	dbsnp151	Pop max freq	Domain	CADD score	VEP	Pph2 status	SIFT_status	LOH	sWGS
EOC_1	chr17:37649046-37649046	C	A	c.C2148A/p.S716R	rs777401578	0.00006486	-	13.78	Moderate	Benign	Tolerated	**ROH**	No TD-Plus
EOC_2	chr17:37687363-37687363	G	A	c.G4267A/p.A1423T	rs201512860	0.0002166	-	20	Moderate	Benign	Tolerated	NA	NA
EOC_3	chr17:37667830-37667830	A	T	c.A2712T/p.E904D	-	0	Kinase	24.9	Moderate	Probably damaging	Deleterious	NA	NA
EOC_4	chr17:37682310-37682310	G	T	c.G3498T/p.Q1166H	-	0	-	20.2	Moderate	Benign	Deleterious	NA	NA
EOC_5	chr17:37627577-37627577	C	G	c.C1489G/p.Q497E	rs766575927	0.00002639	-	19.9	Moderate	Benign	Tolerated	NA	NA
EOC_6	chr17:37618415-37618417	AAC	-	c.92_94del/p.31_32del	rs780413687	0.00003517	-	-	Moderate	-	-	NA	NA
EOC_7	chr17:37687511-37687511	A	G	c.A4415G/p.Y1472C	rs373240630	0.00049 (AFR)	-	23.8	Moderate	Benign	Deleterious	**ROH**	No TD-Plus
EOC_8	chr17:37682202-37682202	C	G	c.C3390G/p.I1130M	rs376340730	0.00006195	-	14.23	Moderate	Benign	Tolerated	NA	NA
EOC_9	chr17:37682501-37682501	A	G	c.A3692G/p.N1231S	rs538854021	0.00005544 (EAS)	-	17.99	Moderate	Benign	Tolerated	NA	NA
EOC_10	chr17:37646920-37646920	C	T	c.C2039T/p.S680F	rs375518105	0.000155	-	28.3	Moderate	Probably damaging	Deleterious	**LOH**	No TD-Plus
EOC_11	chr17:37686935-37686935	C	T	c.C3839T/p.P1280L	rs148965508	0.006415 (AFR)	-	23.6	Moderate	Benign	Deleterious	**ROH**	No TD-Plus

Ref: reference allele; Alt: alternative allele; Pop max freq: maximum allele frequency in any human population (Non-Finish European – NFE - by default); AFR: African population; EAS: East-Asian; EOC: Epithelial Ovarian Carcinoma; LOH: Loss Of Heterozygosity; ROH: Retention Of Heterozygosity; NA: Not Available; sWGS: shallow Whole Genome Sequencing; TD-Plus: tandem duplication-plus genome profile, characteristic of *CDK12* inactivation.