Supplementary Appendix

This appendix has been provided by the authors to give readers additional information about their work.

Supplement to: Anglesio MS, Papadopoulos N, Ayhan A, et al. Cancer-associated mutations in endometriosis without cancer. N Engl J Med 2017;376:1835-48. DOI: 10.1056/NEJMoa1614814

SUPPLEMENTAL APPENDIX

Cancer associated mutations in non-cancer associated endometriosis

Michael S. Anglesio, Ph.D., Nickolas Papadopoulos, Ph.D., Ayse Ayhan, M.D., Ph.D., Tayyebeh M. Nazeran, M.D., Michaël Noë, M.D., Hugo M. Horlings, M.D., Ph.D., Amy Lum, B.Sc., Siân Jones, Ph.D., Janine Senz, B.Sc., Tamer Seckin, M.D., Julie Ho, M.H.A., Ren-Chin Wu, M.D., Vivian Lac, B.Sc., Hiroshi Ogawa, M.D., Basile Tessier-Cloutier, M.D., Rami Alhassan, M.D., Amy Wang, Yuxuan Wang, B.Sc., Joshua D Cohen, B.Sc., Fontayne Wong, B.Sc., Adnan Hasanovic, M.D., Natasha Orr, B.Sc., Ming Zhang, Ph.D., Maria Popoli, B.Sc., Wyatt McMahon, Ph.D., Laura D. Wood, M.D., Ph.D., Austin Mattox, B.Sc., Catherine Allaire, M.D., James Segars, M.D., Christina Williams, M.D., Cristian Tomasetti, Ph.D., Niki Boyd, Ph.D., Kenneth W. Kinzler, Ph.D., C. Blake Gilks, M.D., Luis Diaz, M.D., Tian-Li Wang, Ph.D., Bert Vogelstein, M.D., Paul J. Yong, M.D., Ph.D., David G. Huntsman, M.D., and Ie-Ming Shih, M.D., Ph.D.

TABLE OF CONTENTS FOR SUPPLEMENTAL APPENDIX:

Supplemental Methods

Patient Specimen Collection Information

Lenox Hill Hospital/ Northwell Health (Hofstra University) Cohort

Seventeen samples were obtained from the Departments of Obstetrics and Gynecology and Pathology from the Lenox Hill Hospital/ Northwell Health (Hofstra University), which is a tertiary center for endometriosis surgery. Cases were identified by searching the pathology databases from both institutions between January 2013 and December 2014. The inclusion criteria were endometriosis containing both epithelial and stromal components by extensive pathology review, lack of cancer or dysplasia associated with the lesions and the size of lesions sufficient for tissue coring.

Seirei Mikatahara Hospital Cohort

Seven tissue samples were obtained from the Department of Pathology at the Seirei Mikatahara Hospital in Japan. As above, the cases were identified by searching the pathology databases from both institutions between January 2013 and December 2014. The inclusion criteria were endometriosis containing both epithelial and stromal components by extensive pathology review, lack of cancer or dysplasia associated with the lesions and the size of lesions sufficient for tissue coring.

British Columbia Women's Centre for Pelvic Pain and Endometriosis Cohort

Patients were seen at the British Columbia Women's Centre for Pelvic Pain and Endometriosis. Deep infiltrating endometriosis lesions were defined clinically as lesions with depths greater than 5 mm and were surgically excised with histopathological confirmation of endometriosis epithelial and stromal cells by pathologists (HMH, CBG, or TMN). Tissues were sourced from the OVCARE Tissue bank (banking performed in accordance with the World Endometriosis Research Foundation Endometriosis Phenome and Biobanking Harmonization Project (EPHect)[1]), with appropriate ethics board review. Although some women had endometriosis lesions that were not classified as deep infiltrating endometriosis (i.e. superficial endometriosis), these lesions were not analysed.

Patients selected for broad-panel based sequencing:

Case 25 is a 40 year-old female post-hysterectomy, with a recurrent deep infiltrating endometriosis nodule at the vaginal vault that was attached to the rectum. After laparoscopic excision, pathological examination of the recurrent nodule revealed polypoid endometriosis. We retrieved the initial nodule from the archives for this patient, however, we were not able to confirm endometriosis in this initial nodule and thus this initial nodule was not evaluated. Only the recurrent nodule was analysed as Case 25.

Case 26 is a 38-year old female, with two deep infiltrating endometriosis nodules of the cul-de-sac that were attached to the rectum and right ovary. At laparoscopy, the two cul-de-sac nodules were excised and processed as

separate samples and both were histologically confirmed as endometriosis. Upon analysis one was mutation positive the other was not (see Table 2).

Case 27 is a 36 year-old female post-hysterectomy with bilateral deep infiltrating endometriosis nodules of the pelvic sidewalls that were attached to the ovaries, but without rectal involvement. Endometriosis was confirmed in both specimens however, tissue sufficient for analysis was recovered from only one nodule.

Laser Capture Microdissection (LCM) and Macrodissection Sample Preparation

For specimens from Lenox Hill Hospital/ Northwell Health (Hofstra University) and/or Seirei Mikatahara Hospital serial 10-µm-thick sections were mounted onto PALM membrane slides, stained with hematoxylin, laser captured using the PALM Laser capture microdissection microscope (Zeiss), and catapulted into a tube cap according to the manufacturer's instructions. An estimated total of 500-1000 highly pure normal-appearing epithelial cells or stromal cells were obtained; a serial hematoxylin-eosin slide was used as guide to identify the areas of interest. DNA extraction was performed on pure microdissected cell samples using QIAamp DNA Micro Kit (Qiagen) following the manufacturer's protocol.

For specimens from the British Columbia Women's Centre for Pelvic Pain and Endometriosis, samples were sectioned at 7-8 µm onto PEN membrane slides (Leica Microsystems Inc.). A serial hematoxylin-eosin slide was used as guide to initially identify the areas of interest. Dissection specimens were then deparaffinized with xylene and stained with 10% dilute haematoxylin and eosin. Pathologists (TMN or HMH) marked either combined epithelium and stroma for sequencing (8-15 sections) or separate epithelium and stroma for droplet digital PCR (ddPCR; 3-5 sections), and subsequently microdissected using the LMD7000 Laser Microdissection system (Leica Microsystems Inc.). Control normal tissue was laser captured from surrounding tissues. Macrodissected tissues were stained as defined above and manually macrodissected from standard glass slides using the tip of an 18-guage needle. DNA from both LCM and macrodissected samples was extracted using the ARCTURUS® PicoPure® DNA Extraction Kit (ThermoFisher Scientific) and quantitated using either a Qubit 2.0 Fluorometer (ThermoFisher Scientific) or qPCR. Germline DNA was extracted from saliva using the Oragene DNA kit (DNA Genotek Inc.).

Cored Tissue Sample Preparation

Areas of interest, including lesions and corresponding normal tissues from the same blocks, were cored (2 mm for smaller lesions and 3 mm for larger lesions) from formalin fixed paraffin embedded (FFPE) blocks. H&E sections were prepared before and after coring to confirm the precision of the sampling and to ensure that at least 60% of the core area contained the lesion and 100% contained the normal tissue (see main document Fig. 1). DNA was purified from core samples of individual endometriosis lesions according to Qiagen (Hilden, Germany) tissue material and protocol.

Tissue Fixation

Both the Lennox Hill and Seirei Mikatahara Hospital Cohorts were sourced from hospital pathology archives. All specimens were processed by fixation in neutral buffered

formalin and embedded in paraffin blocks using standard techniques.

Specimens from the BC Women's Centre for Pelvic Pain and Endometriosis were processed as follows: for germline DNA, saliva was pre-operatively obtained from patients using Oragene DNA collection kits (DNA Genotek); for endometriosis specimens, deep infiltrating nodules were excised, intra-operatively, and fixed in Tissue-Tek molecular fixative (Sakura), processed and embedded using the Tissue-Tek microwave rapid processing system (Sakura)[2]. In some cases normal tissue was sampled from molecular fixed and/or formalin fixed tissue as control specimens for ddPCR. These are noted in the text and supplemental tables as originating from the same block as endometriosis specimens. For example, in Case 28, normal endometrium and endo-cervical epithelium were sampled from FFPE tissues. All FFPE tissues were fixed in neutral buffered formalin and paraffin embedded at the Vancouver General Hospital Department of Anatomical Pathology using standard methods.

Whole Exome Capture Sequencing and Copy Number Analysis

DNA was purified from core samples of individual endometriosis lesions according to Qiagen (Hilden, Germany) tissue material and protocol. Libraries were constructed using Tru-Seq reagents from Illumina according to standard protocol provided by Illumina. Exonic regions were captured in solution using the Agilent SureSelect v.5 kit according to the manufacturer's instructions (Agilent, Santa Clara, CA) and modifications as published elsewhere[3]. Next-generation sequencing and bioinformatic analyses of tumor and normal samples were performed either at Johns Hopkins or at Personal Genome Diagnostics (PGDx, Baltimore, MD). Paired-end sequencing, was performed using a HiSeq 2500 Genome Analyzer (Illumina, San Diego, CA). The sequences were aligned to the human genome reference sequence (hg19) using the Eland algorithm of CASAVA 1.8 software (Illumina, San Diego, CA). The chastity filter of the BaseCall software of Illumina was used to select sequence reads for subsequent analysis. The ELANDv2 algorithm of CASAVA 1.8 software (Illumina,San Diego, CA) was then applied to identify point mutations and small insertions and deletions. Known polymorphisms recorded in dbSNP were removed from the analysis. Potential somatic mutations were filtered and visually inspected. Copy number alterations were identified by comparing normalized average per-base coverage for a particular gene in a tumor sample to the normalized average per-base coverage in the matched normal sample.

SafeSeqS

In the Safe-SeqS assay, cored tissue or LCM-derived DNA was aliquoted into 3 wells of a 96-well plate so that an average of 1 to 5 ng of DNA was contained in each well. The DNA from each well was then amplified (15 cycles) using primers containing unique identifier sequences (UIDs), which consisted of 14 random bases with an equal probability of A, C, T, and G, to allow for the distinction of each template molecule. The amplified reactions were purified with AMPure XP beads (Beckman Coulter) and eluted in 250 ml of Buffer EB (Qiagen). One percent (2.5 ul) of purified PCR product was then amplified in a second round of PCR with universal primers, as previously described[4]. The PCR products were purified with AMPure and sequenced on an Illumina MiSeq instrument. High-quality sequence reads were selected on the basis of quality scores, which were generated by the Illumina sequencing instrument to indicate the probability

that an error was made in base calling. The template-specific portion of the reads was matched to reference sequences. Reads from a common template molecule were then grouped on the basis of the UIDs that were incorporated as molecular barcodes. Artifactual mutations introduced during the sample preparation or sequencing steps were reduced by requiring a mutation to be present in >90% of reads in a UID family in order for that UID to be scored as a "supermutant." Wells with fewer than 200 UIDs as a result of poor amplification were excluded. DNA from matched normal tissue were used as controls to call somatic mutations.

The sequence of the primers for the amplification of the KRAS and PIK3CA genomic regions with the mutations reported are:

KRAS
5' TGCTGAAAATGACTGAATATAAACTTGT 3'
5" NNNNNNNNNNNNNNCGTCAAGGCACTCTTGCC 3'

PIK3CA
5' TGAGCAAGAGGCTTTGGAGT 3'
5' NNNNNNNNNNNNNNCCAATCCATTTTTGTTGTCCA 3'

Ns indicate the UID.

Targeted Sequencing

Panel based sequencing using the TruSeq Amplicon Cancer Panel (TSACP; Illumina Inc.) and a previously described TruSeq Custom Amplicon Panel[5] was performed using the TruSeq Amplicon Sequencing Low-Input (beta and v2) library kit followed by MiSeq sequencing (see Table S1 for manifest detail). Libraries were constructed using 12-25 ng of DNA from endometriosis specimens or 50 ng of DNA from saliva in duplicate, barcoded, and run using Illumina 600 cycle MiSeq v3 cartridges. Variants were identified in BaseSpace using the MiSeq Reporter Somatic Variant Caller (Illumina Inc.) tool. Only variants absent from germline DNA and present in both libraries (technical replicates) and both panels (i.e. in quadruplicate) were prioritized for orthogonal validation by ddPCR (see also Figure S1).

The manifest from the Illumina TruSeq Amplicon Cancer Panel (TSACP) can be found on the Illumina website under catalogue number 15032433B (http://support.illumina.com/downloads/truseq_amplicon_-_cancer_panel_manifest_file.html)

Droplet Digital (ddPCR) Details

KRAS

A 10-cycle pre-amplification was performed and diluted pre-amplified DNA was analysed by ddPCR using the RainDance RainDrop system (RainDance Technologies). Custom TaqMan probes (ThermoFisher Scientific) for the c.35G>A (G12D) KRAS variant, or custom LNA PrimeTime® probes for c.35G>T (G12V) and c.35G>C (G12A) variants (Integrated DNA Technologies, Inc.) were used in ddPCR assays (see Supplemental

Methods). Droplets were quantitated using the RainDance Sense instrument and results analysed using the RainDance Analyst V2 software (RainDance Technologies). Controls were included in each run to designate wildtype, mutant, and negative cluster regions. No template control and multiple normal tissue laser captured controls were also included to assess contamination and background probe signal.

The primer sets used for PCR amplification and pre-amplification were as follows:

5'-GCCTGCTGAAAATGACTGAATATAAACT -3'
(forward; Applied Biosystems, Inc.)

5'-GCTGTATCGTCAAGGCACTCTT -3
(reverse; Applied Biosystems, Inc.).

The following probe sets were used in 5-prime exonuclease ddPCR assays:

G12D (c.35G>A) assay
Wildtype 5'-VIC-TTGGAGCTGGTGGCGTA-NFQ-3'
Mutant 5'-FAM-TTGGAGCTGATGGCGTA-NFQ-3'

G12V (c.35G>T) assay
Wildtype 5'-Yak Yellow-CGCC+A+C+CA+GCT-IABkFQ-3'
Mutant 5'-FAM-CG+CC+A+A+CAGC+TC-IABkFQ-3'

G12A (c.35G>C) assay
Wildtype 5'-Yak Yellow-CGCC+A+C+CA+GCT-IABkFQ-3
Mutant 5'-FAM- CGCC+A+G+CA+GCT-IABkFQ-3'

G12R (c.34G>C) assay
Wildtype 5'-VIC-TTGGAGCTGGTGGCGTA-NFQ-3'
Mutant 5'-FAM-TTGGAGCTCGTGGCGTA-NFQ-3'

G12C (C.34G>T) assay
Wildtype 5'-VIC-TTGGAGCTGGTGGCGTA-NFQ-3'
Mutant 5'-FAM-TTGGAGCTTGTGGCGTA-NFQ-3'

DNA pre-amplification was carried out over 10 cycles (polymerase activation of 95 ^{o}C for 10 minutes then 10 cycles of 95^{o}C for 15 seconds and 4 minutes at 60 ^{o}C).
Droplet generation was performed in a 25uL reaction, consisting of 1x ABI genotyping master mix, 1x droplet stabilizer, 1x primer/probe mix for G12D, G12R and G12C variants or 500nM primer+ 200 nM probe mix for G12A and G12V variants, and diluted pre-amplification PCR product.
PCR cycle amplification for ddPCR analysis were as follows: initial polymerase activation for 10 minutes at 95 ^{o}C, then 45 cycles of 95 ^{o}C for 15 seconds, 60 ^{o}C for 45 seconds (with slow ramp of 0.5^{o}C per second to reach temperature), and final denaturation of 98 ^{o}C for 10 minutes.

All individual assays were controlled using sheared DNA (wildtype genomic DNA (Promega Corp.) and cell lines (HEY for G12D; OVCAR5 for G12V, H2009 for G12A, PK-8 for G12R, and MIA PaCa-2 for G12C). To establish the background of our ddPCR

assays we ran multiple additional control specimens from normal (fixed) tissues and non-mutant cell lines. Following these empirical data we chose a negative-cutoff threshold for mutation detection of 5X the maximum allelic frequency obtained from laser captured negative control samples, which is slightly more conservative than the manufacturers recommended *minimum* standard of 3X the negative control.

Empirically determined positive threshold for each assay are as follows:

G12D (c.35G>A) assay > 1.87%
G12V (c.35G>T) assay > 1.53%
G12A (c.35G>C) assay > 1.61%
G12R (c.34G>C) assay *undetermined (only positive control specimens recorded mutant droplet counts)*
G12C (C.34G>T) assay > 0.165%

PPP2R1A

The c.767C>T (S256F) PPP2R1A mutation-status was evaluated using the pre-designed ddPCR BioRad assay (unique assay ID: dHsaMDS850509105). The ddPCR was performed using the QX200 platform (BioRad). After generating 40ul of droplets with the QX200 Droplet Generator, the droplets were transferred into PCR plates. The PCR amplification was performed using following conditions: initial enzyme activation at 95 ^{o}C for 10 minutes, then 40 cycles of 95 ^{o}C for 30 seconds, 55 ^{o}C for 1 minute (with slow ramp of 2 ^{o}C per second to reach temperature), and final enzyme deactivation at 98 ^{o}C for 10 minutes. Fluorescence signal was recorded for each droplet (FAM for mutation and Hex for wild-type) on QX200 Droplet Reader. Results were analyzed using QuantaSoft version 1.7.4 (BioRad). To assess contamination, no template controls (NTC) containing water, were included.

ARID1A immunohistochemistry

ARID1A immunoreactivity was used as a surrogate for ARID1A inactivating mutation [6-8]. FFPE tissue sections were stained manually. Antigen retrieval was performed by placing sections in citrate buffer (pH 6.0) and autoclaving for 10 min. The specificity of the polyclonal rabbit anti-ARID1A antibody (Sigma-Aldrich, HPA005456,) was confirmed in a previous study[9]. The sections were incubated with the primary antibody at a dilution of 1:2,000 and subsequently incubated with the appropriate secondary antibodies. A positive reaction was detected using the EnVision1System (Dako, Carpinteria, CA). Tumor stromal cells served as positive internal controls for ARID1A.

Figure S1

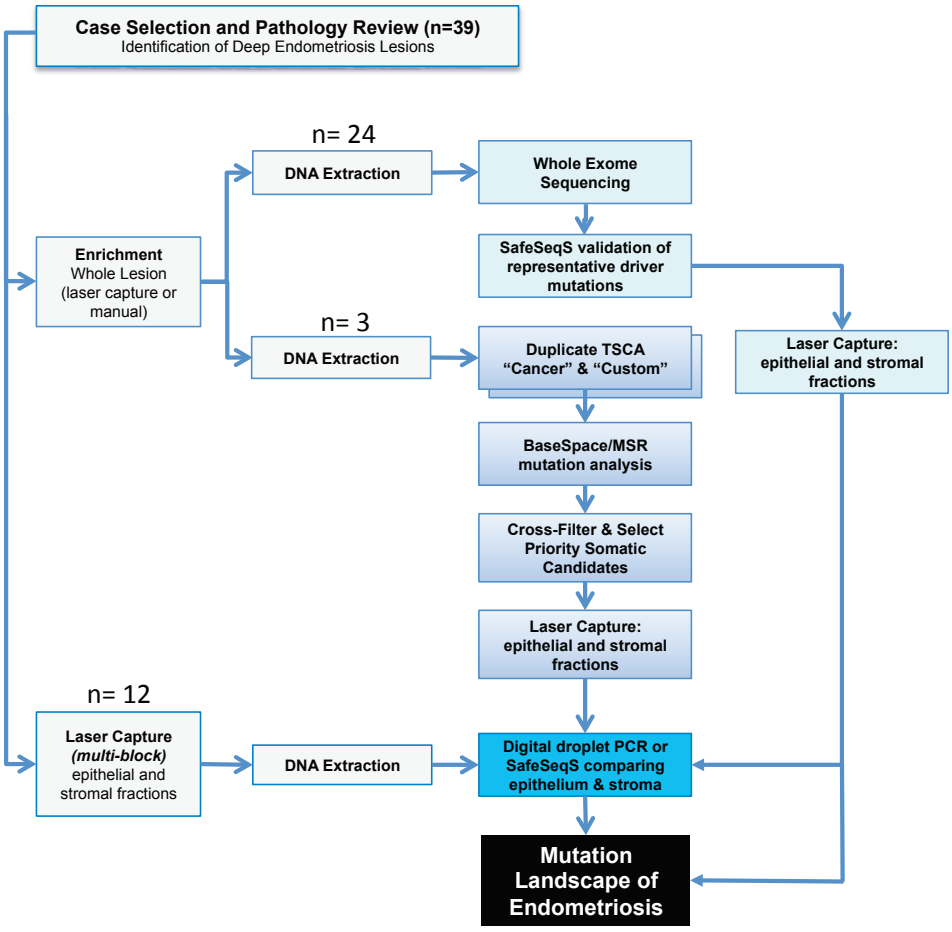
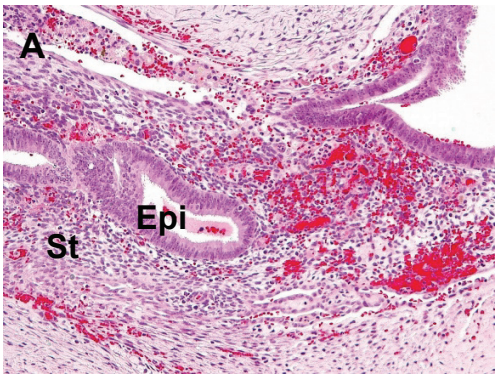
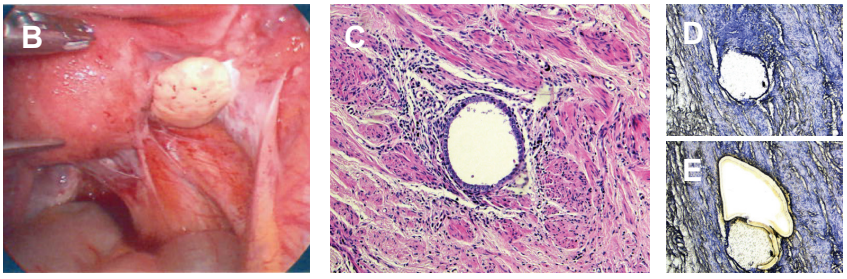


Figure S1: Comprehensive workflow for exome-wide, targeted, and droplet digital PCR analysis including orthogonal validation steps (see also methods).

Figure S2



Case 15 *KRAS* c.35G>A (p.G12V)	
Exome Seq	SafeSeqS
6.55% (lesion)	**59.6%** (epithelium)
0.04% (normal)	0.00% (stroma)



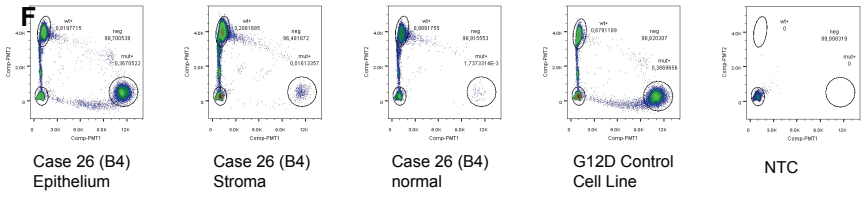


Figure S2: laser-capture and independent digital genomics methods confirm epithelial-exclusive cancer-driver gene mutations. (A) Photomicrograph H&E of deep infiltrating endometriosis illustrating the regions of glandular epithelium and stroma captured by microdissection, lower table show the results from SafeSeqS digital genomic analysis confirming the presence of an activating somatic KRAS mutation confined to the glandular epithelium, consistent with low allelic frequency observed in the initial exome sequencing. (B) Surgical image of the deep endometriosis nodule in Case 26 (C) photomicrograph H&E of a single gland of endometriosis with surrounding stroma used to identify endometriosis from Case 26. (D-E) Lighty H&E stained sections of the endometriosis lesion pre- and post laser dissection of separate epithelial and stromal cells. (F) ddPCR plots showing a high density of fluorescent droplets corresponding to mutation-positive DNA fragments from Case 26 epithelium (Allele frequency 30.93%) but only background level signal in stroma (0.50%), normal tissue from the same tissue block (0.17%) is also show along with a cell line and no-template control (NTC).

List and Description of Supplemental Tables

(All supplemental tables can be found in the accompanying excel file)

Table S1

Supporting statistics from whole-exome sequencing analysis

Table S2

MAF file of all predicted variants from whole-exome sequencing

Table S3

DNA copy number estimation of genes with altered copy number from whole exome sequencing

Table S4

Summary of SafeSeq-S validation analysis

Table S5 Custom Panel Manifest

Contains a manifest of genomic regions used for the Illumina TruSeq Custom Assay (TSCA). The complimentary manifest from the Illumina TruSeq Amplicon Cancer Panel (TSACP) can be found on the Illumina website under catalogue number 15032433B (http://support.illumina.com/downloads/truseq_amplicon_-_cancer_panel_manifest_file.html)

Table S6 Somatic-Variants

Contains a list of ALL variants found from both custom TSCA and TSACP sequencing panels from fixed (MFPE/FFPE) specimens. Data has been filtered only to remove variants found in matched germline samples. Low frequency, low quality, and variants outside of the regions of overlap between custom TSCA and TSACP panels have NOT been removed. Caution should be exercised when examining mutations without orthogonal validation.

Table S7 ddPCR-Raw-Data

Contains a list of all ddPCR assays and runs, including controls, on LCM and macrodissected tissues. KRAS assays for G12V, G12C, G12R, G12A, G12C are noted above with conditions and sequences.

REFERENCES

1. Fassbender A, Rahmioglu N, Vitonis AF, et al. World Endometriosis Research Foundation Endometriosis Phenome and Biobanking Harmonisation Project: IV. Tissue collection, processing, and storage in endometriosis research. Fertility and sterility 2014;102:1244-53.

2. Turashvili G, Yang W, McKinney S, et al. Nucleic acid quantity and quality from paraffin blocks: defining optimal fixation, processing and DNA/RNA extraction techniques. Exp Mol Pathol 2012;92:33-43.

3. Jiao Y, Shi C, Edil BH, et al. DAXX/ATRX, MEN1, and mTOR pathway genes are frequently altered in pancreatic neuroendocrine tumors. Science 2011;331:1199-203.

4. Kinde I, Wu J, Papadopoulos N, Kinzler KW, Vogelstein B. Detection and quantification of rare mutations with massively parallel sequencing. Proceedings of the National Academy of Sciences of the United States of America 2011;108:9530-5.
5. Anglesio MS, Wang YK, Maassen M, et al. Synchronous Endometrial and Ovarian Carcinomas: Evidence of Clonality. Journal of the National Cancer Institute 2016;108.
6. Ayhan A, Mao TL, Seckin T, et al. Loss of ARID1A expression is an early molecular event in tumor progression from ovarian endometriotic cyst to clear cell and endometrioid carcinoma. Int J Gyn Cancer 2012;22:1310-5.
7. Guan B, Mao TL, Panuganti PK, et al. Mutation and loss of expression of ARID1A in uterine low-grade endometrioid carcinoma. Am J Surg Pathol 2011;35:625-32.
8. Wu RC, Ayhan A, Maeda D, et al. Frequent somatic mutations of the telomerase reverse transcriptase promoter in ovarian clear cell carcinoma but not in other major types of gynaecological malignancy. The Journal of pathology 2014;232:473-81.
9. Mao TL, Ardighieri L, Ayhan A, et al. Loss of ARID1A Expression Correlates With Stages of Tumor Progression in Uterine Endometrioid Carcinoma. Am J Surg Pathol 2013;37:1342-8.