



**Kaiser Foundation Health Plan
of Washington**

**Clinical Review Criteria
Genetic Screening and Testing**

Kaiser Permanente Clinical Review Criteria are developed to assist in administering plan benefits. These criteria neither offer medical advice nor guarantee coverage. Kaiser Permanente reserves the exclusive right to modify, revoke, suspend or change any or all of these Review Criteria, at Kaiser Permanente's sole discretion, at any time, with or without notice. **Member contracts differ in their benefits. Always consult the patient's Medical Coverage Agreement or call Kaiser Permanente Customer Service to determine coverage for a specific medical service.**

Criteria

For Medicare Members

Palmetto GBA is the Medicare contractor for Molecular Diagnostic Testing – this site has the most up to date Medicare coverage guidelines for genetic testing.

[Palmetto GBA](#)

Local Coverage Decisions (LCD)

L36198	MoIDX- CDD: NSCLC, Comprehensive Genomic Profile Testing	81445, 81455, 81479
L36362	MoIDX: Biomarkers in Cardiovascular Risk Assessment	82172, 82610, 83090, 83695, 83700, 83704, 86719, 86141
L36163	MoIDX: BRCA1 and BRCA2 Genetic Testing	81162, 81211, 81212, 81213, 81214, 81215, 81216, 81217, 81432, 81445, 81455, 81479
L36386	MoIDX: Breast Cancer Assay: Prosigna	0008M
L36316	MoIDX: Breast Cancer IndexSM Genetic Assay	81479
L36312	MoIDX: CYP2C19, CYP2D6, CYP2C9, and VKORC1 Genetic Testing	81225, 81226, 81227, 81355
L36325	MoIDX: GeneSight® Assay for Refractory Depression	81479
L36186	MoIDX: Genetic Testing for BCR-ABL Negative Myeloproliferative Disease	81206, 81207, 81208, 81219, 81270, 81402, 81403, 81445, 81450, 81455, 81479
L36159	MoIDX: Genetic Testing for Hypercoagulability / Thrombophilia (Factor V Leiden, Factor II Prothrombin, and MTHFR)	81240, 81241, 81291
L36374	MoIDX: Genetic Testing for Lynch Syndrome	81210, 81288, 81292, 81293, 81294, 81295, 81296, 81297, 81298, 81299, 81300, 81301, 81317, 81318, 81319, 81403, 81435, 81479
L36149	MoIDX: HLA-B*15:02 Genetic Testing	81381
L36192	MoIDX: MGMT Promoter Methylation Analysis	81287
L36186	MoIDX: Genetic Testing for BCR-ABL Negative	81206, 81207, 81208, 81219,
L36544	MoIDX: HLA-DQB1*06:02 Testing for Narcolepsy (L36544)	81383

L36256	MoIDX: Molecular Diagnostic Tests (MDT)	0001M, 0002M, 0003M, 0004M, 0006M, 0007M, 0008M, 81161, 81201, 81202, 81203, 81206, 81207, 81208, 81210-81217, 81225, 81226, 81235, 81240, 81241, 81245, 81246, 81256, 81261, 81262, 81263, 81264, 81265, 81270, 81275, 81287, 81288, 81291, 81293, 81294, 81296, 81297, 81298, 81299, 81300, 81301, 81310, 81315, 81316, 81317, 81318, 81321,
		81322, 81323, 81332, 81340, 81341, 81342, 81370-81406, 81479, 85999, 86849, 87999, 88199, 88299, 88380, 88381, 88399, 89398, G0452
L36171	MoIDX: Molecular RBC Phenotyping	81403
L36329	MoIDX-CDD: ConfirmMDx Epigenetic Molecular Assay	81479
L36345	MoIDX-CDD: Decipher® Prostate Cancer Classifier Assay	81479
L36350	MoIDX-CDD: Prolaris Prostate Cancer Genomic Assay	81479
L36335	MoIDX –NRAS Genetic Testing	81311, 81479
L36557	MoIDX: Chromosome 1p/19q Deletion Analysis [PDF]	81402, 88367, 88368, 88369, 88373
L36947	MoIDX – CDD: Oncotype DX Breast Cancer for DCIS (Genomic Health) (L36947)	81479
L36368	Mol-DX-CDD Genomic Health ONCOTYPE DX Prostate Cancer Assay	81479
L36891	MoIDX-CDD: Percepta© Bronchial Genomic Classifier	81479

For Non-Medicare Members

Members must meet all of the following criteria:

1. The member is at clinical risk for a genetic condition because of current documented symptoms being displayed or a strong family history of the condition.
2. The test is scientifically valid and can be adequately interpreted.
3. The results will directly affect a member’s clinical management or reproductive decisions.
4. After appropriate clinical work-up, and informed consent by the appropriate practitioner, the genetic test is indicated.

Genetic testing is not covered for the medical management of a family member who does not have Kaiser Permanente coverage.

For **specific tests listed** below the member must meet the criteria above **AND** the specific test criteria below:

Cardiology	Criteria
Arrhythmogenic Right Ventricular Cardiomyopathy-	MCG* (GHC-0627)
Brugada Syndrome - CACNA1C, CACNA2D1, CACNB2, GPD1L, HCN4, KCND3, KCNE3, KCNE5, KCNJ8, RANGRF, SCN1B, SCN3B, SCN5A, SCN10A, and TRPM4 Genes	MCG*(GHC-0594)
Catecholaminergic Polymorphic Ventricular Tachycardia - RYR2 and CASQ2 Genes	MCG* (GHC-0636)
Coronary Artery Disease - 9p21 Allele	MCG* (A-0657)
Coronary Artery Disease - KIF6 Gene	MCG*(A-0656)
Coronary Artery Disease Gene Expression Testing	MCG* (A-0652)

Cardiology	Criteria
Coronary Artery Disease Genetic Panel	MCG* (A-0658)
Dilated Cardiomyopathy-ANKRD1, DMD, LDB3, LMNA, MYBPC3, MYH7, RBM20, TNN13, TNNT2 Genes	MCG* (GHC-0648)
Hypertrophic Cardiomyopathy-Sarcomere Genes	MCG* (GHC-0633)
Long QT Interval Syndromes-Cardiac Ion Channel Genes	No additional criteria need to be met beyond numbers 1 - 4 on page one.
Thoracic Aortic Aneurysm and Aortic Dissection (Hereditary) - Gene Panels	A-0911 This is not covered per MCG*

Endocrinology	Criteria
Diabetes Mellitus - ABCC8, CEL, FTO, GCK, HNF1A, HNF1B, HNF4A, INS, KCNJ11, KLF11, NEUROD1, and PDX1 Genes	There is insufficient evidence in the published medical literature to show clinical utility.
Diabetes Mellitus, Type 2 - KCNJ11, KCNQ1, PPARG, and TCF7L2 Genes	A-0826: This is not covered per MCG*

Gastroenterology	Criteria
HLA Testing for Celiac Disease:	<ol style="list-style-type: none"> 1. Is medically appropriate for symptomatic patients <ol style="list-style-type: none"> a. Despite being on a gluten free diet OR b. With indeterminate serology/biopsy results 2. It is not covered for <ol style="list-style-type: none"> a. Asymptomatic people OR b. Screening
Hemochromatosis - HFE Gene	MCG* (GHC-0599)
Pancreatitis, Hereditary - PRSS1 Gene	MCG* (GHC-0646)

Genomic Testing Methods and Technologies	Criteria
Cytochrome P450 - 3A4/3A5 Genotyping	A-0775: This is not covered per MCG*
Chromosomal Microarray Analysis (CMA) - Congenital Anomalies	A-0917: This is not covered per MCG*
Micro Array for Evaluation of Intellectual Disability	<ol style="list-style-type: none"> 1) Chromosomal microarray testing may be considered medically necessary for genetic evaluation of an individual when all of the following criteria are met: <ol style="list-style-type: none"> a) There is documentation in the medical record of significant dysmorphic physical features or congenital anomalies b) Results are expected to affect clinical management of the patient for whom the test is ordered c) Testing has been requested following evaluation and genetic counseling by a healthcare professional who has appropriate genetic training and experience. 2) Chromosomal microarray testing may be considered medically necessary for prenatal diagnosis using genetic material obtained by chorionic villus sampling (CVS) or amniocentesis when: <ol style="list-style-type: none"> a) There is a known chromosomal deletion or duplication identified in at least one biologic parent or a biologic sibling, or b) A fetal malformation known to be potentially associated with a

Genomic Testing Methods and Technologies	Criteria
	<p>chromosomal abnormality has been identified by fetal ultrasound and diagnosis by fluorescence in situ hybridization (FISH) testing alone is not possible, and</p> <p>c) Results are expected to directly affect reproductive decisions for the parent(s) or clinical management of the fetus or newborn.</p> <p>3) Chromosomal microarray testing may be considered medically necessary for testing of one or both parents when a chromosomal deletion or duplication has been identified in one or more of their offspring and:</p> <p>a) Parental testing is necessary to guide a reproductive decision, or</p> <p>b) Parental testing is necessary to determine the clinical significance of the chromosome abnormality found in the child, and</p> <p>c) The result is expected to directly affect clinical management of the child</p> <p>The following are not covered:</p> <ul style="list-style-type: none"> • CMA testing for the evaluation of developmental disability, intellectual disability, or autism spectrum disorders in the absence of significant physical dysmorphism. • CMA to confirm the diagnosis of a disorder or syndrome that is routinely diagnosed based on clinical evaluation alone.
Genome-Wide Association Studies	A-0531: This is not covered per MCG*
MicroRNA Detection	A-0705: This is not covered per MCG*
MicroRNA Detection - Heart Failure	A-0838: This is not covered per MCG*
MicroRNA Detection - Inflammatory Bowel Disease	A-0839: This is not covered per MCG*
MicroRNA Detection - Ischemic Heart Disease	A-0840: This is not covered per MCG*
MicroRNA Detection - Kidney Disease	A-0841: This is not covered per MCG*
Molecular Profiling	A-0789: This is not covered per MCG*
Noninvasive Prenatal Testing (Cell-Free Fetal DNA) - Fetal Rhesus D (RhD) Genotyping	A-0847: This is not covered per MCG*
Noninvasive Prenatal Testing (Cell-Free Fetal DNA) - Microdeletion Syndromes	A-0848: This is not covered per MCG*
Noninvasive Prenatal Testing (Cell-Free Fetal DNA) - Monogenic Disorders	A-0849: This is not covered per MCG*
Noninvasive Prenatal Testing (Cell-Free Fetal DNA) - Sex Chromosome Disorders	A-0850: This is not covered per MCG*
Septin 9 (SEPT9) DNA Methylation Testion	A-0706: This is not covered per MCG*
Telomere Analysis	A-0672: This is not covered per MCG*
Topographic Genotypic- PathFinder TG	A-0632: This is not covered per MCG*
Whole Genome/Exome Sequencing - Immunodeficiency Disorders	A-0866: This is not covered per MCG*
Whole Genome/Exome Sequencing - Intellectual Disability	A-0867: This is not covered per MCG*
Whole Genome/Exome Sequencing - Metabolic Disorders	A-0868: This is not covered per MCG*
Whole Genome/Exome Sequencing - Mitochondrial Disorders	A-0869: This is not covered per MCG*

Genomic Testing Methods and Technologies	Criteria
Whole Genome/Exome Sequencing - Autism Spectrum Disorders	A-0870: This is not covered per MCG*
Whole Genome/Exome Sequencing - Neurologic Disorders	A-0871: This is not covered per MCG*
Whole Genome/Exome Sequencing - Congenital Anomalies	A-0872: This is not covered per MCG*
Whole Genome/Exome Sequencing	No additional criteria need to be met beyond numbers 1 - 4 on page one.

Hematology	Criteria
Alpha Thalassemia - HBA1 and HBA2 Genes	MCG* (A-0808)
Beta Thalassemia - HBB Gene	MCG* (A-0815)
Fetal and Neonatal Alloimmune Thrombocytopenia - Human Platelet Antigen (HPA) Genotyping	MCG* (A-0793)
Factor V Leiden Thrombophilia-F5 gene	No longer require medical review
Fanconi Anemia-FANC Gene	MCG* (GHC-0683)
Hemoglobinopathies, Thalassemias and Sickle Cell Disease - HBA1, HBA2, and HBB Genes	MCG* (GHC-0604)
Hyperhomocysteinemia - MTHFR Gene	MCG*(A-0629)
Post-Transfusion Purpura - Human Platelet Antigen (HPA) Genotyping	A-0800: This is not covered per MCG*

Hematology	Criteria
Prothrombin Thrombophilia - F2 Gene	MCG* (GHC-0613)
Sickle Cell Disease - HBB Gene	MCG* (A-0864)
Von Willebrand Disease-VWF Gene	MCG* (GHC-0688)

Metabolic and Developmental Disorders	Criteria
Angelman Syndrome DNA Methylation Testing - UBE3A Gene	MCG* (GHC-0708)
Ashkenazi Jewish Genetic Panel	MCG*(GHC-0592)
Autism Spectrum Disorders – Gene Panels	A-0914 This is not covered per MCG*
Beckwith-Wiedemann Syndrome - CDKN1C Gene	MCG* (A-0765)
Bloom Syndrome-BLM Gene	MCG*(GHC-0682)
Canavan Disease - ASPA Gene	MCG* (GHC-0595)
Deafness, Hereditary - GJB2 and GJB6 Genes	MCG* (GHC-0596)
Developmental Delay - Gene Panels	A-0925 This is not covered per MCG*
Fragile X-Related Disorders-FMR1 Gene	MCG* (A-0602)
Fragile X-Associated Primary Ovarian Insufficiency - FMR1 Gene	MCG* (A-0829)
Fragile X-Associated Tremor/Ataxia Syndrome - FMR1 Gene	MCG* (A-0830)
Gaucher Disease - GBA Gene	MCG* (GHC-0603)
Glycogen Storage Disease, Type 1 G6PC Gene	MCG* (GHC-0684)
Intellectual Disability - Gene Panels	A-0923 This is not covered per MCG*
Joubert Syndrome - AHI1, ARL13B, CC2D2A, CEP41, CEP290, CSPP1, C5orf42, INPP5E, KIF7, NPHP1, OFD1, RPGRIP1L, TCTN1, TCTN2, TCTN3, TMEM67 (MKS3), TMEM138, TMEM216, TMEM231, and TMEM237 Genes	MCG* (A-0785)
Lesch-Nyhan Syndrome - HPRT1 Gene	MCG* (GHC-0606)
Maple Syrup Urine Disease-BCKDHA	MCG* (GHC-0681)
Maple Syrup Urine Disease, Type 3 - DLD Gene	MCG* (A-0776)

Metabolic and Developmental Disorders	Criteria
Mucopolysaccharidosis IV-MCOLN1 Gene	MCG* (GHC-0686)
Niemann-Pick Disease (Acid Sphingomyelinase Deficiency) - SMPD1, NPC1, and NPC2 Genes	MCG* (GHC-0611)
Prader-Willi Syndrome DNA Methylation Testing	MCG* (GHC-0707)
Rett Syndrome-FOXP1 and MECP2 Genes	MCG* (GHC-0687)
Tay-Sachs Disease and Variants - HEXA Gene	MCG* (GHC-0614)
Usher Syndrome - ADGRV1 (GPR98), CDH23, CIB2, CLRN1, DFNB31, HARS, MYO7A, PCDH15, USH1C, USH1G, and USH2A Genes	MCG* (A-0802)

Miscellaneous	Criteria
Autosomal and X-Linked Recessive Disease Carrier Screening - Expanded Gene Panels	A-0768: This is not covered per MCG*
Familial Mediterranean Fever-MEFV Gene	MCG* (GHC-0689)
Hereditary Hemorrhagic Telangiectasia - ACVRL1, ENG, and SMAD4 Genes	MCG* (GHC-0704)
Male Infertility - Y Chromosome Microdeletion Analysis	MCG* (A-0803)
Malignant Hyperthermia Susceptibility-RYR1 Gene	MCG* (GHC-0690)

Nephrology	Criteria
Polycystic Kidney Disease (Autosomal Recessive) - PKHD1 Gene	MCG* (A-0852)

Neurology	Criteria
Alzheimer Disease - APOE, APP, CLU, CR1, PICALM, PSEN1, and PSEN2 Genes	MCG* (GHC-0590)
Alzheimer Disease (Late Onset) - APOE Genotyping	A-0809: This is not covered per MCG*
Amyotrophic Lateral Sclerosis (ALS) - SOD1 Gene	No additional criteria need to be met beyond numbers 1 - 4 on page one.
Ataxia-Telangiectasia - ATM Gene	MCG* (GHC-0593)
CADASIL (Cerebral Autosomal Dominant Arteriopathy with Subcortical Infarcts and Leukoencephalopathy) - NOTCH3 Gene	MCG* (GHC-0668)
Charcot-Marie-Tooth Hereditary Neuropathy, Type 1 - EGR2, FBLN5, LITAF, MPZ, NEFL, and PMP22 Genes	MCG* (A-0691)
Charcot-Marie-Tooth Hereditary Neuropathy, Type 2 - MFN2 Gene	MCG* (A-0816)
Charcot-Marie-Tooth Hereditary Neuropathy, Type 4 - GDAP1 and PRX Genes	MCG* (A-0818)
Charcot-Marie-Tooth Hereditary Neuropathy, Type X - GJB1 and PRPS1 Genes	MCG* (A-0819)
Familial Dysautonomia- IKBKAP Gene	MCG* (GHC-0685)
Familial Frontotemporal Dementia - C9orf72, GRN, and MAPT Genes	MCG* (A-0906) This is not covered per MCG
Huntington Disease - HTT Gene	MCG* (GHC-0605)
Muscular Dystrophies (Duchenne, Becker) - DMD Gene	MCG* (GHC-0608)
Myotonic Dystrophy-DMPK Gene	MCG* (GHC-0609)
Myotonic Dystrophy, Type 2 - CNBP Gene	MCG* (A-0844)

Nemaline Myopathy - ACTA1, CFL2, KBTBD13, KLHL40, KLHL41, LMOD3, NEB, TNNT1, TPM2, and TPM3 Genes	MCG* (A-0792)
Parkinson Disease - ATP13A2, LRRK2, PARK2, PARK7, PINK1, and SNCA Genes	MCG* (A-0671)
Seizure Disorders (Hereditary) - Gene Panels	MCG* (A-0905) This is not covered per MCG
Spinal Muscular Atrophy - SMN1 and SMN2 Genes	MCG* (GHC-0659)

Oncology	Criteria
Acute Lymphoblastic Leukemia - BCR-ABL1 Fusion Gene Testing	MCG* (A-0759)
Acute Promyelocytic Leukemia - PML-RARA Fusion Gene Testing	MCG* (A-0760)
BRACAnalysis Large Rearrangement Test (BART)	MCG* (GHC-0638)
Breast Cancer Gene Expression Assays	See Oncotype Dx
Breast Cancer - HER2 Testing	MCG* (A-0766)
Breast or Ovarian Cancer, Hereditary - BRCA1 and BRCA2 Genes (see genetic testing and screening policy)	MCG* (GHC-0499)
Cancer of Unknown Primary: Gene Expression Profiling	MCG* (A-0673)
Chronic Eosinophilic Leukemia/Hypereosinophilic Syndrome - FIP1L1-PDGFRΑ Fusion Gene Testing	MCG* (A-0770)
Chronic Myelogenous Leukemia - BCR-ABL1 Fusion Gene Testing	MCG* (A-0771)
Cologuard	See Fecal DNA Testing
Colon Cancer Gene Expression Assay - ColoPrint	A-0822: This is not covered per MCG*
Colon Cancer Gene Expression Assay - GeneFx Colon	A-0821: This is not covered per MCG*
Colorectal Cancer - KRAS and NRAS Genes	MCG* (A-0773)
Colorectal Cancer - BRAF V600E Testing	Colorectal Cancer - BRAF V600E Testing BRAF V600E testing may be indicated for one of the following: <ol style="list-style-type: none"> 1. Metastatic colorectal cancer and therapy with cetuximab and pantitumumab is being considered OR 2. All of the following: <ul style="list-style-type: none"> o Colorectal cancer (newly diagnosed and untreated) o Need to screen for Lynch syndrome, as indicated by 1 or more of the following: <ul style="list-style-type: none"> ▪ Tumor immunochemistry abnormal for MLH1 protein expression ▪ Tumor with microsatellite instability (MSI-high)
Cowden Syndrome - PTEN Gene	MCG* (GHC-0585)
DecisionDx- Melanoma	There is insufficient evidence in the published medical literature to show clinical utility.
Familial Adenomatous Polyposis-APC Gene	MCG* (A-0534)
MUTYH-Associated Polyposis - MUTYH Gene	MCG* (A-0828)
Familial Adenomatous Polyposis - Gene Panels	A-0827: This is not covered per MCG*
Gastric Cancer, Hereditary - CDH1 Gene	MCG* (A-0779)
Gastrointestinal Stromal Tumor (GIST) - KIT and PDGFRA Genes	MCG* (A-0780)

Oncology	Criteria
Glioblastoma Multiforme - Gene Expression Profiling (DecisionDx-GBM)	A-0781: This is not covered per MCG*
Gynecologic Cancer (Hereditary) - Gene Panel	A-0782: This is not covered per MCG*
Li-Fraumeni Syndrome - TP53 Gene	MCG* (GHC-0584)
Lymphoma - T-Cell Antigen Receptor (TCR) Gene Rearrangement Testing	MCG* (A-0786)
Lynch Syndrome - MLH1, MSH2, MSH6, and PMS2 Genes (see genetic panel policy)	MCG* (GHC-0533)
Malignant Melanoma, Familial - CDK4 and CDKN2A Genes	No additional criteria need to be met beyond numbers 1 - 4 on page one.
Malignant Melanoma (Uveal) - BAP1, CDK4, and CDKN2A Genes	A-0836: This is not covered per MCG*
Malignant Melanoma - BRAF V600 Testing	MCG* (A-0787)
Melanoma (Cutaneous) - Gene Expression Profiling	A-0837: This is not covered per MCG*
Melanoma Gene Expression Profiling	MCG* (A-0670)
Multiple Endocrine Neoplasia (MEN) Syndromes - MEN1 Gene	MCG* (GHC-0582)
Multiple Endocrine Neoplasia (MEN) Syndrome, Type 2 - RET Gene	MCG* (A-0842)
Myelodysplastic Syndromes - Gene Panels	A-0791: This is not covered per MCG*
Myeloproliferative Neoplasms - JAK2 Genes	MCG* (A-0669)
Myeloproliferative Neoplasms - MPL Gene	MCG* (A-0843)
Neuroblastoma - ALK, MYCN, and PHOX2B Genes and Gene Expression Profiling	MCG* (A-0610)
Neurofibromatosis - NF1 Gene	MCG* (A-0581)
Neurofibromatosis - NF2 Gene	MCG* (A-0846)
Non-Small Cell Lung Cancer - KRAS Gene Testing	A-0851: This is not covered per MCG*
Non-Small Cell Lung Cancer - Anaplastic Lymphoma Kinase (ALK) Fusion Gene Testing	MCG* (A-0794) <i>criteria on hold for approval</i>
Non-Small Cell Lung Cancer - EGFR Gene Testing	MCG* (A-0795) <i>criteria on hold for approval</i>
OVA1- Assessment for Ovarian Cancer	There is insufficient evidence in the published medical literature to show clinical utility.
Paraganglioma-Pheochromocytoma Syndromes, Hereditary - SDHB, SDHC, SDHD, and TMEM127 Genes	MCG* (GHC-0535)
Paraganglioma-Pheochromocytoma (Hereditary) - Gene Panel	A-0798: This is not covered per MCG*
Pancreatic Cancer (Hereditary) - Gene Panel	A-0797: This is not covered per MCG*
Prostate Cancer - Genetic Profiles, BRCA2, MMR, PCA3, PTEN, and TMPRSS-ETS Fusion Genes	MCG* (A-0612)
Prostate Cancer Gene Expression Testing - Oncotype DX	A-0712: This is not covered per MCG*
Prostate Cancer - HOXB13, MMR, PTEN, and TMPRSS2-ETS Fusion Genes	A-0854: This is not covered per MCG*
Prostate Cancer - PCA3 Gene	A-0855: This is not covered per MCG*
Prostate Cancer Gene Expression Testing - Decipher	A-0856: This is not covered per MCG*
Prostate Cancer Gene Expression Testing - Prolaris	A-0857: This is not covered per MCG*
Proteomics - Ovarian Cancer Biomarker Panel (ROMA)	A-0858: This is not covered per MCG*
Proteomics (VeriStrat)	Epidermal Growth Factor Receptor Testing is covered when: 1) Diagnosis of NSCLC

Oncology	Criteria
Peutz-Jeghers Syndrome - STK11 Gene	MCG* (A-0799)
Renal Cancer (Hereditary) - Gene Panel	A-0801: This is not covered per MCG*
Retinoblastoma - RB1 Gene	MCG* (GHC-0586)
Thyroid Nodule Gene Expression Testing (Afirma)	There is insufficient evidence in the published medical literature to show clinical utility.
Von Hippel-Lindau Syndrome - VHL Gene	MCG* (GHC-0583)
Wilms Tumor - WT1, WT2, and WTX Genes	MCG* (GHC-0615)

Ophthalmology	Criteria
Age-Related Macular Degeneration - Gene Panels	A-0913 This is not covered per MCG
Retinal Disorders - Gene Panels	A-0912 This is not covered per MCG

Orthopedics	Criteria
Adolescent Idiopathic Scoliosis - ScoliScore	A-0761: This is not covered per MCG*
Ankylosing Spondylitis - HLA-B27 Testing	MCG* (A-0762)
Osteogenesis Imperfecta - COL1A1, COL1A2, CRTAP, FKBP10, IFITM5, LEPRE1, PLOD2, PMB1, PPIB, SERPINF1, SERPINH1, SP7, SPARC, TMEM38B, and WNT1 Genes	MCG* (A-0796)

Pulmonary	Criteria
Asthma - ADRB2 Gene	A-0763: This is not covered per MCG*
Cystic Fibrosis-CFTR Gene and Mutation Panel:	MCG* (GHC-0597)

Risk Prognosticator Test	Criteria
<ul style="list-style-type: none"> BREVAGen™ Fibroblast Growth Factor Receptor 3 (FGFR3) OVA1™ Test for the Assessment of Suspected Ovarian Cancer MammaPrint Test (Gene-Expression Profiling Test, 70-Gene Prognostic Signature) 	There is insufficient evidence in the published medical literature to show that this service/therapy is as safe as standard services/therapies and/or provides better long-term outcomes than current standard services/therapies.

*The MCG are proprietary and cannot be published and/or distributed. However, on an individual member basis, Kaiser Permanente can share a copy of the specific criteria document used to make a utilization management decision. If one of your patients is being reviewed using these criteria, you may request a copy of the criteria by calling the Kaiser Permanente Clinical Review staff at 1-800-289-1363.

If requesting this service, please send the following documentation to support medical necessity:

- Any genetic counseling notes if applicable Results of prior genetic testing
- Last 6 months of specialist notes of that is being reviewed (i.e., neurological notes, medical oncology notes, cardiology notes)

The following information was used in the development of this document and is provided as background only. It is not to be used as coverage criteria. Please only refer to the criteria listed above for coverage determinations.

Evidence and Source Documents

[Afirma](#)

[Array Comparative Genomic Hybridization \(aCGH\) for the Genetic Evaluation of Patients with Intellectual Disability DecisionDx- Melanoma](#)

[HLA Testing for Celiac Disease](#)

[Micro Array for Evaluation of Intellectual Disability](#)

[OVA1 for Assessment of Ovarian Cancer](#)

[Risk Prognosticator Tests](#)

[Thyroid Nodule Gene Expression Testing \(Afirma\)](#)

Background

Genetic screening is used to identify the genetic disorders or the potential for transmission of genetic disorders in populations at risk for a particular genetic disorder. Genetic screening is only appropriate when the natural history of the disease is understood; the screening tests are valid and reliable; sensitivity, specificity, false-negative, and false-positive rates are acceptable; and effective therapy is available. A sufficient benefit must be derived from a screening program to justify its cost.

Medical Technology Assessment Committee(MTAC)

Afirma

BACKGROUND

Thyroid nodules are clinically identified in 5-7% of the population, and incidentally on ultrasonography in up to 50% of women and 20% of men over the age of 50. Thyroid nodules are typically benign, but 5-15% prove to be malignant. It is thus recommended that any identified nodule measuring one centimeter or more in diameter be diagnostically evaluated. Ultrasound-guided fine needle aspiration (FNA) biopsy is the most widely used method for clinical evaluation of a suspicious thyroid nodule. FNA is a safe and simple outpatient procedure that yields cellular material suitable for cytological analysis. It can identify approximately 50% of malignant nodules and 70% of benign nodules without the need to perform a diagnostic surgery. However, 15-30% of the biopsied nodules have indeterminate cytology and cannot be conclusively diagnosed by FNA biopsy alone. Most patients with indeterminate lesions or lesions suspicious for malignancy, according to the Bethesda classification* system, are referred to surgery for both diagnostic and therapeutic purposes. Surgery is the recommended and appropriate treatment for thyroid cancer, however 70-75% of the nodules with indeterminate FNA cytology are found to be benign on final surgical pathology. Thus, a large proportion of these patients may undergo unnecessary partial or complete thyroidectomy with its potential surgical complications and risk of long-term morbidity (Alexander 2012, Duick 2012, Walsh 2012, Ali 2013, Labourier 2015, Sacks 2016).

Molecular markers and assays have been investigated for their ability to preoperatively classify the indeterminate thyroid nodules. Each has its performance characteristics and diagnostic values. Ideally a molecular marker or panel of markers is accurate in differentiating benign from malignant in any lesion that is considered suspicious or indeterminate. Molecular tests should also be simple to use, reproducible by the different institutions/laboratories, and cost-effective.

Molecular genetic testing for cytologically indeterminate thyroid nodules fall in two approaches: the “rule in” and the “rule-out” disease approach. Tests that rule-in malignancy (such as BRAF, RAS mutations, RET/PTC and PAX8-PPAry) have high specificity and positive predictive values (PPVs) for malignancy by identifying specific mutations or gene rearrangements known to be present in thyroid cancer. However, they have limited sensitivity and negative predictive values (NPVs) and fail to detect as many as 30% of malignancies. Tests that rule-out the disease on the other hand, should have a high sensitivity and negative predictive value in order to exclude malignancy when the test results are benign. Because a majority of nodules with indeterminate cytology are found to be benign on surgical resection, a test that can preoperatively rule-out malignancy may spare a subset of these patients unnecessary diagnostic surgeries (Alexander 2012, Kouniavsky 2012, Ward 2013, Chaudhary 2016, Nishino 2016).

*2008 Bethesda classification system for thyroid cytology: Class I: Nondiagnostic or unsatisfactory, Class II: Benign, Class III: atypia or follicular lesion of undetermined significance (AUS/FLUS), Class IV: follicular neoplasm or suspicious for follicular neoplasm (FN), Class V: suspicious for malignancy (SUSP) and Class VI: malignant) (Kuo, 2016)

Afirma gene expression classifier (GEC) is a molecular test developed by Veracyte Inc. (San Francisco, CA) with the intention of reducing unnecessary diagnostic surgeries in patients with thyroid nodules with indeterminate FNA cytopathologic results. It represents the “rule-out” approach by preoperatively identifying the benign thyroid nodules and ruling-out malignancy. Afirma GEC uses a proprietary diagnostic algorithm that analyses the mRNA expression of 167 genes to identify the signature of benign thyroid nodules. 142 of the 167 genes are in the main classifier, and 25 genes filter out rare neoplasms. The selected gene profile is based on the gene expression identified from FNAs of surgically proven benign and malignant thyroid nodules. During the Afirma GEC test RNA is extracted from the FNA sample, amplified and hybridized to a custom microarray to examine for gene patterns. These are compared with the GEC proprietary panel, which molecularly classifies them as either ‘benign’ or ‘suspicious’. Insufficient RNA in the sample leads to ‘no result’ conclusion in approximately 10% of cases. Nodules with benign results, in addition to clinical judgement, are typically followed up clinically and ultrasonography, while those with suspicious results undergo diagnostic thyroid lobectomy with possible total thyroidectomy (Alexander 2012, Kim 2012, Ward 2013,

Kuo 2016, Witt 2016).

Afirma GEC is a proprietary test commercially owned by Veracyte Corporation and is offered through a sole source, which is a Clinical Laboratory Improvement Amendments certified [CLIA] reference laboratory. During a routine FNA of a thyroid nodule, after the aspirates are obtained for cytopathologic examination, two more needle passes are obtained for Afirma analysis and immediately stored in a preservative. These are either 1. Sent to a Veracyte independent industry partner (Thyroid Cytopathology Partners [TCP], Austin, TX) that performs cytopathologic exam of the FNA sample, and only runs the Afirma test for indeterminate diagnoses on cytopathology, or 2. In Thyroid Cytopathology Medical centers designated as “Enabled centers” cytopathology is done in-house and specimens with indeterminate results based on the Bethesda criteria are sent for Afirma GEC testing. Afirma test is run only on nodules with indeterminate cytology. If the cytopathologic evaluation reveals any other diagnosis or is nondiagnostic due to insufficient FNA samples, the preserved samples are discarded. The goal of the test is to identify the benign nodules from among those with indeterminate cytopathology. It is not intended to assist with clinical decision making for patients who have an indication for surgery or meet criteria for surgical interventions (Alexander 2012, Duick 2012, Ward 2013, Kuo 2016, Yip 2016).

03/20/2017: MTAC REVIEW

Evidence Conclusion: Analytic validity of Afirma GEC (From an earlier MTAC review)

Evaluating the analytic performance of Afirma GEC includes studying the stability of RNA in FNAs during collection, storage, and shipment; reproducibility of the test; and its analytic sensitivity and specificity under various conditions e.g. interference of the assay with bloody FNA and genomic DNA. The literature search revealed one study (Walsh and colleagues, 2012) that evaluated the analytic performance of Afirma GEC in a number of sub-studies. The investigators obtained prospective FNA samples aspirated in vivo from 43 patients from outpatient clinics, preoperatively, or immediately after surgical excision. The samples were placed in FNAProtect preservative solution and shipped chilled or frozen, then stored at -80°C upon receipt. The RNA was extracted and its yield examined for quantity and quality using positive (tissue lysate) and negative (water) as controls. Three different lots of controls were tested over several weeks of independent runs by 3 different operators to determine reliability of the test. Multiple lots of benign and malignant total RNA were manufactured and used as process controls to determine the analytic sensitivity of the test using different RNA input quantity and under different dilution of malignant FNA content. These studies indicated tolerance to variation in RNA input across a range of 5-25 ng, as well as dilution of malignant FNA material down to 20%. Analytic specificity of the test using malignant samples mixed with blood (up to 83%) and genomic DNA (up to 30%) showed minimal assay interference. However benign FNA samples mixed with relatively high proportions of blood had a potential for yielding false positive results. The investigators examined the stability of RNA in FNAs during collection and shipment and found that RNA content within FNAs preserved in FNAProtect was stable for up to 6 hours at room temperature with no change in RNA yield, and that the FNA storage and shipping temperatures had no significant effect on GEC scores. They also examined the reproducibility of the test and indicated that it was reproducible from extraction through GEC results, including variation across operators, runs, reagent lots, and laboratories. The investigators concluded that the analytical performance and reproducibility of the Afirma Gene Expression Classifier was successfully verified. The research was supported by Veracyte Corporation (the maker of Afirma GEC), and the authors of the study had financial ties to the corporation.

Clinical validity of Afirma GEC An ideal diagnostic test would have high sensitivity and specificity to correctly detect or exclude a condition. A molecular diagnostic test with high sensitivity offers a high negative predictive value (NPV) when the risk of malignancy is low, and can “rule out” malignancy. Conversely, a test with high specificity offers high positive predictive (PPV) value and can “rule in” malignancy. A preoperative diagnostic test would ideally have a high sensitivity in order not to miss a malignant nodule, and have a high NPV to avoid surgery in patients with benign nodules. Predictive values do not only depend on the sensitivity and specificity of the test, but also on the prevalence of the disease; e.g. as the disease prevalence increases, the NPV decreases and vice versa. Afirma GEC test was validated in a. a double-blind prospective multicenter study (Alexander 2012 (Evidence table 1, reviewed earlier by MTAC). The study involved 265 nodules with indeterminate cytology that were selected for GEC analysis. Molecular results were compared to the gold standard of post-surgical histopathology. The malignancy rate was 32% and the Afirma sensitivity and specificity were 92% and 52% respectively with a NPV of 94-95% and PPV of 27-38% for Bethesda III/IV nodules. In the subgroup in patients with nodules suspicious for malignancy (SUSP) the NPV was only 85%, based on which, the authors concluded that GEC should not be used for cytology SUSP nodules. The study was conducted to validate the GEC accuracy by comparing it to surgical histopathology, and did not compare its performance to repeat FNA or other immunochemical or molecular testing. A number of post-validation analyses were conducted by independent or industry supported investigators. In the initial validation study (Alexander 2012) the decision to resect the nodule was made independently of the GEC test results, while in the post-validation studies GEC was a factor influencing the decision whether to recommend a diagnostic surgical

intervention. The published studies and analyses showed a wide variation in the NPVs and PPVs of Afirma GEC test results. The NPV and PPV of a test are neither absolute nor inherent in the test, but depend on the pre-test probability of malignancy in the population studied, i.e. prevalence of malignancy in indeterminate nodules. This is made clear by Marti and colleagues (2014) who retrospectively analyzed all indeterminate thyroid nodules (ITNs) evaluated with GEC at two centers with widely different prevalence of malignancy in ITNs (Memorial Sloan Kettering Cancer Center [MSK] and Mount Sinai Beth Israel [MSBI]) (see table below).

The results of the validation study as well as post-validation studies are summarized in the following table.

Performance of GEC in the validation study and selected post-validation studies

Study	N FNA ITN undergoing GEC	Afirma results Suspicious	Benign	Inadequate RNA/ no result	Cancer prevalence in indeterminate FNA	NPV‡	PPV‡‡	Operative rate in GEC benign results
Alexander 2012 (multicenter Validation study)	265	62%	38%		32%	94-95%	27-38%	NA
Alexander, 2014 (5 centers)***	339	40%	55%	5%	--	99.4%	--	6.3%*
Harrell, 2014 (One practice)	58	62%	35%	3%	33-36%	88.3-89.6%	56%	25%**
Mclver, 2014	72	61%	22%	17%	17%	94%	15.6%	25%
Marti, 2015 MSK center(tertiary)	94	74%	26%	--	30-38%	86-92%	57.1%	8%
MSBI comprehensive health system)	71	48%	52%	--	10-19%	95-98%	14.3%	14%
Chaudhary, 2016†	158	54%	40%	6%		100%	38%	13%
Witt, 2016 (single-practice)	32	47%	44%	9%	21%	100%	--	0
Samulski, 2016 (single institution)	294	46%	49%	5%		81% for resected nodule, 98% for unrested benign GEC	39%	12%
Sacks, 2016 (single medical center)	140	55.7%	37.1%	7.1%	31.5%	92%††	33.3%	

All studies were performed in Institutional Enabled Centers, except for Harrell (2014) study where the cytology specimens were sent to Thyroid cytopathology partners (TCP).

* 1/11 (9.1%) was malignant (false negative)

** Of the 20 benign GEC patients 5 underwent surgery 2 of which (40%) were found to be malignant (False negative),

*** There were variation between the 5 study sites in the cytology distribution and Afirma GEC performance. GEC suspicious cases proved to be cancerous in 44% of cases (False positive in 56% of cases)

‡ The NPVs (the probability of cancer in GEC benign nodules) were all estimates and could not be directly assessed because not all patients had undergone surgery to determine surgical pathology or had long-term follow-up of the GEC benign nodules.

‡‡ The reported PPVs ranged between studies from 14-57% which limits the utility of the test as a rule-in test i.e. to predict the risk of malignancy.

† A comparison between pre- and post-GEC era showed no significant difference in surgical excision rates of FNA ITN. There were differences in the accuracy and predictive values of the GEC according to the cytomorphological features of the nodules. The authors concluded that the GEC test was found to reduce surgical excision of nodules with suspicious for follicular neoplasm (SFN), but not with FLUS / AUS or Hurthle cell neoplasm (HCN). They recommended repeating FNA rather than performing Afirma GEC test for FLUS/AUS, and be cautious when ordering GEC on HCN cases. 8 (13%) of the benign affirm underwent surgery and all were found benign

†† Estimated based on the prevalence. Could not be calculated due o the low number of GEC-benign cases with surgical pathology

-- Not provided

Based on the results of the published studies, some investigators suggest that Afirma GEC may provide useful information in practice settings where the prevalence of malignancy in indeterminate thyroid nodules is 15-21%. At this range, and using the sensitivity and specificity data from the multicenter validation study the NPV would be >95% and the PPV >25%. It is suggested that GEC may also provide some useful information with the prevalence of malignancy ranging from 12-25%, but is not expected to be useful in altering management if the prevalence is outside this range (Marti 2015, Zhang 2016). The Afirma GEC performance was found to be suboptimal for Hurthle cell neoplasms (HCNs). Wu and colleagues (2016) examined the clinical factors influencing the performance of GEC testing and found that the test has a limited clinical validity for HCNs due to the high rate of false positive results (specificity 22.7-26.1% and PPV 29.2%). Other studies also showed inconsistent and low performance of GEC testing for HCN nodules. In the clinical validation study only 4 of 21 (19%) FNA samples from Hurthle adenomas were classified as benign with GEC.

Limitaons in the published studies These include but are not limited to the following:

- All analyses were retrospective with potential bias and confounding.
- There were intra- and inter-observer differences within and between studies in the histological interpretations.
- Only data for patients with GEC testing were analyzed and with the exception of one study, the results were not compared to repeat FNA or other tests.
- The NPVs were all estimates as only a very limited number of GEC benign nodules underwent surgery, and the follow-up duration was too short to determine the true benign nature of the GEC benign nodules.
- The majority of the published studies were industry sponsored.
- The predictive values of a test vary with the prevalence of the disease in a population studied and may not be generalized to other groups. A better analysis would be the likelihood ratios which are not affected by prevalence.

Santhanam and colleagues' meta-analysis (2013, Evidence table 2) pooled the results of 7 prospective and retrospective studies to determine the sensitivity and specificity of the GEC test in classifying FNA indeterminate thyroid nodules, and evaluate its clinical utility. The results of the meta-analysis are summarized in the following table:

Pooled results

Pooled values	Value (95% CI)	P value
Sensitivity	95.7% (92.2-97.9)	0.09
Specificity	30.5% (26.0-35.3)	<0.01
Positive likelihood ratio*	1.20 (0.99-1.44)	<0.01
Negative likelihood ratio**	0.2 (0.11-0.36)	0.56
Diagnostic odds ratio	7.86 (4.1-15.01)	0.42
Prevalence of malignancy	37.1%	
Positive predictive value	44.8 (40.4-49.4)	

*A good test for ruling-in a disease is the one with the largest positive likelihood ratio (LR) A positive LR of 1 means that the test does not provide any information on ruling in the disease, LR >1<5 indicates a small effect, and LR>10 indicates a large effect on increasing the probability of a disease is presence.

** The better test to rule-out a disease is the one with the smaller negative likelihood ratio. LR <0.1 indicates that the result has a large effect on decreasing the probability of the disease (rule out), LR 0.1-0.5 indicates moderate effect, and >0.5 indicates a small effect. The meta-analysis had valid methodology, but a meta-analysis is as good as the studies it includes. Due to the lack of RCTs and comparative prospective studies Santhanam and colleagues pooled the results of observational prospective and retrospective studies. There was significant heterogeneity between the studies as they were performed at different institutions and included a wide distribution of patients with different indeterminate cytology results (the test may perform better for one type of neoplasm/cancer versus the other). The meta-analysis had the advantage of calculating likelihood ratios which are not affected by prevalence the condition as the predictive values. However likelihood ratios are calculated based on the sensitivity and specificity of the test, which may have not been accurate as the majority of GEC benign cases did not undergo surgery or were followed up for a sufficient duration to assess the actual accuracy of the test, and not all GEC suspicious cases underwent surgery. More recently in 2016, an international panel of pathologists and clinicians reclassified a clinically indolent malignant tumor (encapsulated follicular variant of papillary thyroid carcinoma [EFVPT]) as a benign neoplasm (noninvasive follicular thyroid neoplasm with papillary-like nuclear features [NIFTP]) (Nikiforov 2016). This reclassification may affect the calculated performance of the current Afirma GEC as it has not been validated with these changes. In one study Samulski and colleagues (2016) reported that of 11 NIFTP cases in their cohort, only one was classified as benign with the GEC test.

Clinical utility of Afirma GEC The clinical utility of Afirma GEC would be guiding the management decisions by clearly ruling out malignancy in FNA indeterminate nodules to avoid unnecessary diagnostic surgery. The published studies on the impact of Afirma GEC on the management of patients with FNS ITNs were retrospective

in nature, performed in different sites with intra- and inter-rater variability, which are potential sources of selection and performance bias. In addition the studies only focused on nodules that underwent Afirma GEC testing, and did not investigate the effect of FNA results on overall thyroidectomy rates, or include a comparison group to examine the impact of a repeat FNA or other tests for nodules with indeterminate cytopathology. [Santhanam and colleagues \(2013, Evidence table 2\)](#) discussed in the previous section on clinical validity of Afirma GEC also evaluated its clinical utility of the test. The authors calculated that for patients with FNA indeterminate nodule, one thyroid surgery can be avoided for every two Afirma GEC tests, assuming that >90% of the patients with benign GEC are followed conservatively. They noted however, that according to the American Cancer Society, the 5-year survival of stage I and stage II follicular and papillary thyroid cancer is 100%. The morbidity and mortality rates in patients with FNA indeterminate thyroid nodules is reported to be more likely low, and thus the diagnosis of suspicious nodules with GEC testing may represent a lead-time bias with little change in overall survival.

[Sacks and colleagues \(2016 Evidence table 3\)](#) performed a retrospective analysis to evaluate the impact of Afirma GEC testing on cytopathology diagnosis, rate of surgery, and the rate of malignancy on all indeterminate nodules (ITNs) before and after the introduction of Afirma GEC testing at a high volume thyroid center. The study was a retrospective analysis of patient data from one institution, with no direct comparison to a control group. However, it had the advantage of reporting on outcomes of repeat FNAs, comparing two cohorts' pre-and post-Afirma, and reporting on thyroidectomy rates among all cases irrespective of GEC testing. The calculated PPV for the test was 33.3%, and the estimated NPV was 92% (an accurate NPV could not be calculated due to small number of GEC benign cases with surgical pathology). There was a significant increase Bethesda III-IV diagnosis in the post Afirma cohort compared to the pre-Afirma cohort (13.4% vs. 10.7%, $p<0.005$), with a corresponding significant decrease in benign cytology (Bethesda II) post-Afirma (74.6% compared to 68.8% pre-Afirma, $p<0.001$), despite the use of the same guidelines, practice, reporting scheme, and personnel. In an attempt to explain the reason for this shift, the authors supposed that cytopathologists, especially those with less experience, may be less likely to classify nodules as Bethesda II knowing that the GEC testing will help stratify them. No significant changes were observed for Bethesda I, V, or VI, or in the rate of repeat FNA for ITNs. The author noted that while Afirma may reduce the rate of thyroidectomy for nodules with benign GEC results, the "suspicious label" may increase it. Only 33.3% of GEC suspicious cases were found to be malignant. The analysis shows that 35.2% of patients with ITNs who underwent a repeat FNA were classified as non-ITN and avoided Afirma testing. Overall, the results of the analysis indicate that the use of Afirma GEC testing was associated with an increase in the rate of FNA indeterminate diagnosis, and a decrease in the incidence of benign diagnosis. GEC testing did not reduce the overall rate of thyroidectomies which is its main goal. As indicated earlier the study had its disadvantages, which may limit generalization of the results.

[Abeykoom and colleagues \(2016\)](#), performed a similar respective analysis in a single endocrine clinic comparing the rate of surgeries pre-and post GEC testing for nodules with indeterminate cytopathology (N=61 [27 before and 34 after GEC implementation]). The results were however, inconsistent with Sack's 'findings. The analysis showed no significant difference before and after GEC implementation in the rate of ITNs, but there was a significant decrease in the recommendation for surgery for patients with ITNs from 81.5% pre-GEC implementation to 50% post GEC ($p=0.01$). The surgical pathology for those who underwent an operation was read as malignant in 20% and 85.7%.of patients before and after Afirma GEC respectively ($p<0.01$). The study was retrospective, small, included patients from a single center over two-time periods with different pathologists analyzing the specimens, which are potential sources of confounding and bias that may limit generalization of the results.

[Duick and colleagues \(2012, Evidence table 4, from an earlier MTAC review\)](#) performed a chart review for 21 endocrinology practices in 11 states. They analyzed data for 368 patients with 395 cytologically indeterminate thyroid nodules with Afirma GEC benign results. 7.6% of these patients underwent surgery and 94.4% were managed nonoperatively.

The study did not have a comparison group, but the authors compared the 7.6% surgical rate in nodules with benign GEC results to a 74% historical rate of diagnostic surgery ($P<0.001$). The main indications for surgery for those with GEC benign results were the rapid growth or larger size of the nodules, local pressure symptoms, or the presence of a second suspicious nodule or malignant nodule.

The study was retrospective, used a historical comparison, and investigated the decision-making of endocrinologists experienced in managing patients with thyroid nodules, which may differ from that made by primary care providers or other specialists. In addition the authors of the study did not provide data on long-term follow-up of those who were managed by watchful waiting rather than surgery.

[Sipos and colleagues \(2016\)](#), retrospectively analyzed data recorded for 98 patients with a benign GEC over a

mean duration of 36 ±3 months (range 0-44 months) treated at multiple centers. 17 of these 98 patients (17.3%) underwent surgery during this period. 88% of the surgeries were performed in the first 2 years after the benign GEC results with the rate leveling after the first year. The most common indications for surgery were the nodules rapid growth and large size. The authors concluded that the study shows that benign GEC test results are associated with low operative rates. The study had its limitations and the authors did not provide data on the pathology results of the resected nodules.

Articles: The updated literature search revealed a number of retrospective analyses performed after the Afirma GEC validation study, a meta-analysis that pooled the results of selected studies, and three retrospective studies on the clinical utility of the test. The study on the analytic validity, the two clinical validation studies as well as two retrospective studies on clinical utility were reviewed earlier by MTAC. The meta-analysis and the more recent studies on the clinical validity and clinical utility of Afirma GEC test were reviewed and their results summarized.

Array Comparative Genomic Hybridization (aCGH) for the Genetic Evaluation of Patients with Intellectual Disability

BACKGROUND

Intellectual disability, also termed mental retardation or cognitive disability, affects approximately 1-3% of the general population and is defined as a significant impairment in cognitive and adaptive functions, with the age of onset before 18 years. It is a serious and lifelong condition that presents significant challenges to families and to public health. Determining the specific etiology of intellectual disability may help to provide answers related to prognosis, recurrence risk, and treatment. Intellectual disability can be caused by anything that damages or interferes with the growth or maturation of the brain; however, genetic (chromosomal) abnormalities are one of the main causes of intellectual disability (Galasso 2010, Sagoo 2009). Chromosomal abnormalities are deletions and duplications of genomic material and are commonly referred to as copy number variations. Conventional methods for detecting these abnormalities include karyotyping and florescent in situ hybridization (FISH). Karyotyping involves visualizing the chromosome for large gains or losses in chromosomal material and is generally the first step in cytogenetic analysis. Karyotyping can detect chromosomal abnormalities such as deletions, duplications, inversions, and translocations across the entire genome; however, it lacks the resolution necessary to detect abnormalities smaller than 3-5 megabases (Mb; 3-5 million base pairs). FISH uses florescent-labeled chromosome-specific probes to detect chromosomal abnormalities. FISH can detect submicroscopic abnormalities and is often used in situations where the karyotype is normal, but there is a high clinical suspicion of a deletion syndrome. However, FISH is a targeted method and requires prior knowledge of the chromosome region(s) of interest to request the appropriate FISH test. Additionally, FISH can only screen a limited number of genomic regions at a time (Bremner 2009, Fruhman 2010, Galasso 2010, Gropman 2010). Array comparative genomic hybridization (aCGH) is a more recent technology used to identify copy number variations by comparing patient DNA with reference DNA. It is currently used as an adjunct to conventional methods. There are two types of aCGH: targeted and whole-genome. Targeted arrays are designed to interrogate areas of the genome with known clinically significant abnormalities. Whole genome arrays provide high resolution coverage of the entire genome. This can lead to the discovery of new copy number variations. Compared to conventional methods, aCGH has a higher resolution and is able to simultaneously detect copy number variations in multiple regions of the genome. Additionally, unlike FISH, knowledge of the chromosome region(s) of interest does not need to be determined in advance because a single array assay detects all genomic variants represented on the array. Array CGH is not without limitations. It cannot detect totally balanced translocations or inversions; it performs suboptimally for polyploidy; and has not been optimized for prenatal diagnosis of point mutations. Because aCGH cannot identify the exact location of a duplicated chromosome, further testing with karyotype or FISH may be necessary. Another limitation is the potential to identify novel copy number variants with unknown clinical significance (Fruhman 2010, Moeschler 2008). Array CGH is a laboratory-developed test and is commercially available from several different laboratories. Laboratory-developed tests are licensed under the Clinical Laboratory Improvement Amendments (CLIA) and do not require clearance from the FDA.

The use of Gene Expression Classifier (Afirma®) does not meet the *Kaiser Permanente Medical Technology Assessment Criteria*.

4/18/2011: MTAC REVIEW

Array Comparative Genomic Hybridization (aCGH) for the Genetic Evaluation of Patients with Intellectual Disability

Evidence Conclusion: Analytic validity The BCBS review identified several studies that evaluated the sensitivity of aCGH. The sensitivity of aCGH testing compared to conventional methods (karyotype and/or FISH) ranged from 73% to 100%. As false-positive rates were inconsistently reported, specificity could not be determined (BCBS 2009). Clinical validity

Articles: No studies were identified that evaluated the impact of conventional methods or aCGH on patient outcomes other than diagnostic yield. Results from the BCBS review suggest that diagnostic yield in patients with intellectual disability ranged from 5 to 16.7%, which represents a significant improvement compared to conventional methods. The number needed to test by aCGH to detect one clinically relevant abnormality ranged from 25 to 6 depending on the diagnostic yield. Limitations of these studies include: different aCGH resolution, patient selection criteria ranged from none to stringent criteria, and three different types of arrays were used (targeted, whole-genome, and those that combined targeted and whole-genome arrays) (BCBS 2009).

Diagnostic yield ¹ of aCGH, karyotype, and FISH			
	aCGH	FISH	karyotype
	4-7%	5-6%	
Diagnostic yield	(In those negative by karyotype and FISH)	(In those negative by karyotype)	3-5%

¹Estimates from Stankiewicz 2007.

Clinical utility The BCBS review included two small studies with a high risk of bias and found that there was insufficient evidence to determine the clinical utility of aCGH testing (BCBS 2009). **Conclusion:** Analytic validity: There is fair evidence that aCGH testing had good sensitivity compared to conventional methods; however, there is insufficient evidence to determine the specificity or reproducibility of this test. **Clinical validity:** There is fair evidence that aCGH increases diagnostic yield over conventional methods; however, this is an intermediate outcome. **Clinical utility:** There is insufficient evidence that patients managed with the genetic test had better outcomes than patients managed without the genetic test.

In 2009, Blue Cross and Blue Shield (BCBS) evaluated the use of aCGH for the genetic evaluation of patients with developmental delay/ mental retardation. Studies were selected for review if they were published after the 2009 review and did not support the BCBS recommendations. No studies were identified that would change the BCBS recommendations. The following review was critically appraised: Blue Cross and Blue Shield Association.

Special report: aCGH for the genetic evaluation of patients with developmental delay/mental retardation or autism spectrum disorder. Assessment Program. Volume 23, No. 10. April 2009.

The use of Array Comparative Genomic Hybridization (aCGH) for the genetic evaluation of patients with intellectual disability does not meet the *Kaiser Permanente Medical Technology Assessment Criteria*.

HLA Testing for Celiac Disease

BACKGROUND

Celiac disease is a chronic, autoimmune disorder that affects approximately 1% of children and adults in the United States. In individuals with celiac disease, the ingestion of gluten proteins found in wheat, rye, and barley lead to an autoimmune reaction that causes small intestine mucosal injury. Damages in the small intestine can cause gastrointestinal symptoms and interfere with the absorption of nutrients from food. This may lead to malnutrition-related problems such as anemia, vitamin deficiencies, osteoporosis, and neurological disorders. A gluten-free diet typically resolves symptoms and can prevent long-term consequences (Tack 2010). There are a variety of tests available to diagnose celiac disease. The gold-standard for diagnosing celiac disease is a small intestine biopsy. However, this test is not a perfect gold-standard as false positive and false-negative results may occur due to interobserver variability, patchy mucosal damage, low-grade histological abnormalities, and technical limitations. Additionally, histological features are not unique to celiac disease. Serum antibody tests are used as an initial screening tool to detect and support the presence of celiac disease and to select which patients should undergo a biopsy. Two of the most sensitive and specific serological tests for diagnosing celiac disease are tests that assess the presence of IgA autoantibodies against the endomysium of connective tissue (EMA) (sensitivity 62-81%, specificity 80-99%) and against tissue transglutaminase (tTGA) (sensitivity 81-88%, specificity 84-99%). While these test are accurate, they are not without limitations. For example, the EMA test correlates with the degree of mucosal damage. As such, the sensitivity of this test is lower in patients with milder cases (higher chance of false negative results). Additionally, false negative results may occur in patients with an IgA deficiency and in patients who are already on a gluten-free diet. In patients with an IgA deficiency, serum IgA testing can be replaced by using IgG assays, which are less sensitive than IgA assays. Another test that can be used to rule out the diagnosis of celiac disease is human leukocyte antigen (HLA) genotyping. It has been reported that approximately 90-95% of patients with celiac disease are carriers of the HLA-DQ2 heterodimer and most of the remaining patients carry the HLA-DQ8 heterodimer. Since virtually all patients with celiac disease carry one of these heterodimers, celiac disease is highly unlikely when both are absent. It has been proposed that using HLA genotyping as an initial screening tool may avoid future concerns about the

condition and eliminate further diagnostic testing. However, HLA typing is not a perfect solution since around 25-40% of the general population carries either HLA-DQ2 or DQ8, of which the majority never develop the disease. Other situations where HLA genotyping may be useful is when the diagnosis of celiac disease is unclear based on serological and/or histological findings. Additionally, HLA genotyping can be performed in patients who are already on a gluten free diet (Tack 2010, Hadithi 2010).

4/18/2011: MTAC REVIEW

HLA Testing for Celiac Disease

Evidence Conclusion: Analytic validity There are a variety of methods used for HLA genotyping. Each of these assays has its advantages and limitations (Monsuur 2008, Lavant 2009). Clinical validity A recent prospective cohort study evaluated the accuracy of serologic tests and HLA-DQ genotyping used alone and in combination for diagnosing celiac disease compared to small intestine biopsy. Results from this study suggest that both tTGA and EMA are sensitive and specific tests for diagnosing celiac disease. HLA-DQ testing was also highly sensitive, but was not as specific as serologic testing. The addition of HLA-DQ genotyping to serum antibody tests did not increase test performance compared to serologic testing alone. Results should be interpreted with caution as only 16 patients were diagnosed with celiac disease (Hadithi 2007). Sensitivity and specificity of serologic testing and HLA-DQ typing for diagnosing celiac disease

	Sensitivity (95% CI)	Specificity (95% CI)
HLA-DQ testing		
HLA-DQ2 or DQ8	100 (79-100)	57 (52-62)
Serologic testing using IgA		
tTGA	81 (54-95.9)	99.1 (97.7-99.7)
EMA	81 (54-95.9)	99.1 (97.7-99.7)
tTGA & EMA	81 (54-95.9)	99.3 (98-99.9)
Both serologic testing & HLA-DQ testing		
tTGA & HLA-DQ	81 (54-95.9)	99.3 (98-99.3)
EMA & HLA-DQ	81 (54-95.9)	99.1(97.7-99.8)
tTGA, EMA, & HLA-DQ	100 (79-100)	99.3 (98-99.9)

Abbreviations: EMA= antiendomysium antibody; tTGA= antitransglutaminase antibody.

Another observational study investigated whether HLA genotyping would be useful to identify first-degree relatives of patients with celiac disease who do not need further screening for celiac disease. Fifty-four families with at least two siblings with celiac disease were selected to participate in the study. In total, 245 (52.5%) first-degree relatives agreed to participate. The diagnosis of celiac disease was based on duodenal biopsy and medical records. Of all of the first-degree relatives, 17.6% (N=43) did not carry any of the celiac disease risk alleles. Of these relatives, only one was diagnosed with celiac disease (Karinen 2010). Clinical utility Because of its low specificity HLA genotyping may not be an ideal initial screening test for diagnosing celiac disease. However, HLA genotyping may be useful in certain situations, such as when the diagnosis of celiac disease is unclear based on serologic and histologic findings and when patients are already on a gluten free diet, to rule out celiac disease. Additionally, as negative serologic or histologic test results do not exclude the development of celiac disease later in life, the use of HLA genotyping in patients who are at increased risk for celiac disease may prevent unnecessary serologic and histologic testing. Conclusion: Analytic validity: There are a variety of methods used for HLA genotyping. Each of these assays has its advantages and limitations. Clinical validity: There is fair evidence that HLA genotyping may be a useful adjunct in the diagnosis of celiac disease as it has a high negative predictive value. Clinical utility: No studies were identified that addressed the clinical utility of HLA genotyping for celiac disease; however, early identification and treatment of the disease can prevent short- and long-term complications.

Articles: Articles were selected for review if they included at least 25 subjects and assessed the accuracy of HLA genotyping compared to the small intestine biopsy. A prospective cohort study was selected for review. The following study was critically appraised: Hadithi M, von Blomberg ME, Crusius BA, et al. Accuracy of serologic tests and HLA-DQ typing for diagnosing celiac disease. *Ann Intern Med* 2007; 147:294-302. See [Evidence Table](#)

The use of HLA testing for celiac disease does meet the *Kaiser Permanente Medical Technology Assessment Criteria*.

Micro Array for Evaluation of Intellectual Disability

BACKGROUND

Intellectual disability, also termed mental retardation or cognitive disability, affects approximately 1-3% of the general population and is defined as a significant impairment in cognitive and adaptive functions, with the age of onset before 18 years. It is a serious and lifelong condition that presents significant challenges to families and to

public health. Determining the specific etiology of intellectual disability may help to provide answers related to prognosis, recurrence risk, and treatment.

Intellectual disability can be caused by anything that damages or interferes with the growth or maturation of the brain; however, genetic (chromosomal) abnormalities are one of the main causes of intellectual disability (Galasso 2010, Sagoo 2009).

Chromosomal abnormalities are deletions and duplications of genomic material and are commonly referred to as copy number variations. Conventional methods for detecting these abnormalities include karyotyping and florescent in situ hybridization (FISH). Karyotyping involves visualizing the chromosome for large gains or losses in chromosomal material and is generally the first step in cytogenetic analysis. Karyotyping can detect chromosomal abnormalities such as deletions, duplications, inversions, and translocations across the entire genome; however, it lacks the resolution necessary to detect abnormalities smaller than 3-5 megabases (Mb; 3-5 million base pairs).

FISH uses florescent-labeled chromosome-specific probes to detect chromosomal abnormalities. FISH can detect submicroscopic abnormalities and is often used in situations where the karyotype is normal, but there is a high clinical suspicion of a deletion syndrome. However, FISH is a targeted method and requires prior knowledge of the chromosome region(s) of interest to request the appropriate FISH test. Additionally, FISH can only screen a limited number of genomic regions at a time (Breman 2009, Fruhman 2010, Galasso 2010, Gropman 2010).

Array comparative genomic hybridization (aCGH) is a more recent technology used to identify copy number variations by comparing patient DNA with reference DNA. It is currently used as an adjunct to conventional methods. There are two types of aCGH: targeted and whole-genome. Targeted arrays are designed to interrogate areas of the genome with known clinically significant abnormalities. Whole genome arrays provide high resolution coverage of the entire genome. This can lead to the discovery of new copy number variations. Compared to conventional methods, aCGH has a higher resolution and is able to simultaneously detect copy number variations in multiple regions of the genome. Additionally, unlike FISH, knowledge of the chromosome region(s) of interest does not need to be determined in advance because a single array assay detects all genomic variants represented on the array. Array CGH is not without limitations. It cannot detect totally balanced translocations or inversions; it performs suboptimally for polyploidy; and has not been optimized for prenatal diagnosis of point mutations. Because aCGH cannot identify the exact location of a duplicated chromosome, further testing with karyotype or FISH may be necessary. Another limitation is the potential to identify novel copy number variants with unknown clinical significance (Fruhman 2010, Moeschler 2008).

Array CGH is a laboratory-developed test and is commercially available from several different laboratories. Laboratory-developed tests are licensed under the Clinical Laboratory Improvement Amendments (CLIA) and do not require clearance from the FDA.

04/18/2011: MTAC REVIEW

Array Comparative Genomic Hybridization (aCGH)

Evidence Conclusion: Analytic validity - The BCBS review identified several studies that evaluated the sensitivity of aCGH. The sensitivity of aCGH testing compared to conventional methods (karyotype and/or FISH) ranged from 73% to 100%. As false-positive rates were inconsistently reported, specificity could not be determined (BCBS 2009). Clinical validity - No studies were identified that evaluated the impact of conventional methods or aCGH on patient outcomes other than diagnostic yield. Results from the BCBS review suggest that diagnostic yield in patients with intellectual disability ranged from 5 to 16.7%, which represents a significant improvement compared to conventional methods. The number needed to test by aCGH to detect one clinically relevant abnormality ranged from 25 to 6 depending on the diagnostic yield. Limitations of these studies include: different aCGH resolution, patient selection criteria ranged from none to stringent criteria, and three different types of arrays were used (targeted, whole-genome, and those that combined targeted and whole-genome arrays) (BCBS 2009).

Diagnostic yield¹ of aCGH, karyotype, and FISH			
	aCGH	FISH	karyotype
Diagnostic yield	4-7% (In those negative by karyotype and FISH)	5-6% (In those negative by karyotype)	3-5%

¹Estimates from Stankiewicz 2007.

Clinical utility- The BCBS review included two small studies with a high risk of bias and found that there

was insufficient evidence to determine the clinical utility of aCGH testing (BCBS 2009).

Conclusion:

1. Analytic validity: There is fair evidence that aCGH testing had good sensitivity compared to conventional methods; however, there is insufficient evidence to determine the specificity or reproducibility of this test.
2. Clinical validity: There is fair evidence that aCGH increases diagnostic yield over conventional methods; however, this is an intermediate outcome.
3. Clinical utility: There is insufficient evidence that patients managed with the genetic test had better outcomes than patients managed without the genetic test.

Articles: In 2009, Blue Cross and Blue Shield (BCBS) evaluated the use of aCGH for the genetic evaluation of patients with developmental delay/ mental retardation. Studies were selected for review if they were published after the 2009 review and did not support the BCBS recommendations. No studies were identified that would change the BCBS recommendations. The following review was critically appraised: Blue Cross and Blue Shield Association.

Special report: aCGH for the genetic evaluation of patients with developmental delay/mental retardation or autism spectrum disorder. Assessment Program. Volume 23, No. 10. April 2009.

The use of Array Comparative Genomic Hybridization (aCGH) for the genetic evaluation of patients with intellectual disabilities does not meet the *Kaiser Permanente Medical Technology Assessment Criteria*.

Risk Prognosticator Test

BREVAGen

BACKGROUND

According to the American Cancer Society, breast cancer is the second leading cause of death in women in the United States after lung cancer. Current methods of assessing breast cancer risk include the Breast Cancer Risk Assessment Tool (BCRAT) otherwise known as the Gail model. This model incorporates individual risk factors such as basic demographic information, reproductive history and medical history. Recent genome wide association studies have identified several single nucleotide polymorphisms (SNPs) associated with an increased risk of breast cancer leading to an additional dimension and understanding of risk (Easton, Pooley et al. 2007; Stacey, Manolescu et al. 2007; Stacey, Manolescu et al. 2008). The BREVAGen™ (Phenogen Sciences, Inc., Charlotte, NC) is a risk stratification test for sporadic breast cancer. Intended for use as an adjunct to the Gail model, the test consists of two parts, the first, a series of questions to determine clinical risk and the second, a buccal swab to analyze specific genetic markers. The latter part of the test, includes a panel of seven SNPs associated with breast cancer risk and does not include either of the BRCA mutations. Ultimately, a patient's risk is calculated by multiplying the product of the individual SNP risks by the Gail model risk. According to the BREVAGen™ website, the test is only suitable for women of European descent aged 35 years or older. No test combining the results of SNP analysis with clinical factors to predict breast cancer risk has been approved or cleared by the U.S. Food and Drug Administration (FDA). BREVAGen™ is offered as a laboratory developed tests and only requires oversight under the Clinical Laboratory Improvement Amendments of 1988 (CLIA). The development and use of this laboratory developed test is restricted to laboratories certified as high complexity under CLIA. Under the current regulatory program, CLIA requires that laboratories demonstrate quality systems which includes validation and proficiency testing.

12/16/2013: MTAC REVIEW

BREVAGen

Evidence Conclusion: *Analytic Validity* Although the seven breast cancer associated SNPs were identified in genome wide association studies (Easton, Pooley et al. 2007; Stacey, Manolescu et al. 2007; Stacey, Manolescu et al. 2008), there are no publications that specifically report on the analytic validity of the BREVAGen™ panel. *Clinical Validity* The BREVAGen™ test was clinically validated in a nested case-control cohort study. The study included 1,664 women from the US Women's Health Initiative (WHI) who developed breast cancer between randomization and study completion and 1,636 age-matched breast cancer-free controls (Mealiffe, Stokowski et al. 2010). Overall, the study suggests that the BREVAGen may add predictive accuracy to the Gail Model, however, the degree of improved risk prediction is modest and the clinical implications are not well established. The study is not generalizable as the included population was limited to postmenopausal white non-Hispanic women. In addition, four of the authors were employees of the developer of the BREVAGen™ during the time of publication and the analyses were relevant to the development of the

product. Ideally, additional studies should be completed to further assess the clinical validity of the combined Gail and SNP risk model. *Clinical Utility* Theoretically, the BREVAGen™ test can provide information that can help guide physicians in making

individualized patient management decisions, such as appropriate counseling, screening regimens and risk reduction strategies, however, there is no published literature to support the clinical utility.

Conclusion: There is no evidence to determine the analytic validity of the BREVAGen™. There is some evidence to suggest that the addition of the BREVAGen™ panel is superior in determining breast cancer risk compared to Gail score alone. There is no evidence to determine the clinical utility of the BREVAGen™.

Articles: A search of PubMed was completed for the period through November 2013 for studies on the accuracy of BREVAGen™ for detecting the absence or presence of certain common genetic variations associated with an increased risk for developing breast cancer. The search strategy used the terms BREVAGen, Breast Cancer Risk Tool, Gail Model, genetic risk, single nucleotide polymorphism, breast cancer, and sporadic with variations. To identify ongoing clinical trials, a search of the National Institute of Health Clinical Trials website was also conducted using the same methodology. Articles were limited to those published in the English language with human subject enrollment. The search was supplemented by an examination of article reference lists in addition to the PubMed related articles function. The literature search for BREVAGen™ revealed one publication that clinically validates the Breast Cancer Risk Model in combination with the genetic and clinical information. The following study was selected for review: Mealiffe ME, Stokowski RP, Rhees BK, et al. Assessment of clinical validity of a breast cancer risk model combining genetic and clinical information. *Journal of the National Cancer Institute*. 2010;102(21):1618-1627. [See Evidence Table](#).

The use of BREVAGen does not meet the *Kaiser Permanente Medical Technology Assessment Criteria*.

Fibroblast Growth Factor Receptor 3 (FGFR3) for Urothelial Carcinoma

BACKGROUND

It is estimated that approximately 70,530 new cases of bladder cancer will be diagnosed in the United States in 2010, and 14,680 will die of the disease (Jemal 2010). The most commonly occurring form of bladder cancer in the United States is urothelial carcinoma (also known as transitional cell carcinoma). The clinical spectrum of urothelial carcinoma can be divided into 3 categories: non-muscle-invasive, muscle-invasive, and metastatic disease. This review will focus on non-muscle-invasive urothelial cancer (NMIUC), which makes up approximately 75-80% of urothelial carcinoma. NMIUC includes stage Ta (noninvasive papillary carcinoma), Tis (carcinoma in situ), and T1 (tumor invades subepithelial connective tissue) tumors. The standard treatment for stage Ta, Tis, and T1 tumors is transurethral resection of bladder tumor (TURBT). Depending on prognosis adjuvant intravesical chemotherapy or immunotherapy may also be considered. However, despite treatment a significant number of patients will develop recurrence within 1 to 2 years of the initial treatment. Because of the high risk of recurrence careful surveillance is required for patients with NMIUC (Chou 2010, Cheng 2011, NCCN 2011, Pollard 2010). Assessing the risk of progression and recurrence is important for planning therapy. The risk for tumor progression and recurrence is estimated using factors such as histological grade, stage, depth of invasion, and extent of disease; however, the ability of these factors to predict clinical outcome is limited (Burger 2008, Cheng 2011, NCCN 2011). Recently, it has been suggested that molecular biomarkers such as fibroblast growth factor receptor 3 (FGFR3) may be useful for predicting clinical outcome and planning therapy. FGFR3 regulates cell growth, differentiation, and angiogenesis. More than 70% of low-grade noninvasive papillary urothelial carcinomas harbor FGFR3 mutations. Studies suggest that urothelial carcinomas that harbor FGFR3 mutations may be associated with improved prognosis (Cheng 2011). The CertNDx molecular grading assay (Predictive Biosciences, Inc.) was designed as a tool to be used in conjunction with clinical and histological parameters to aid in the clinical management of NMIUC. This test uses two biomarkers to determine molecular grade. The first biomarker is FGFR3 and the second is Ki-67, which is a marker of cell proliferation (Cheng 2011). Patients with molecular grade 1 (mG1) have FGFR3 mutations and low Ki-67 levels. Patients with molecular grade 2 (mG2) have FGFR3 mutations with high Ki-67 levels or wild-type FGFR3 and low Ki-67 levels. Patients with molecular grade 3 (mG3) are FGFR3 wild-type and have high Ki-67 levels. Patients with molecular grade 1 have favorable prognosis, patients with molecular grade 2 have intermediate prognosis, and patients with molecular grade 3 have poor prognosis.

10/17/2011: MTAC REVIEW

Fibroblast Growth Factor Receptor 3 (FGFR3) for Urothelial Carcinoma

Evidence Conclusion: Analytic validity- No studies were identified that addressed the analytic validity of the CertNDx molecular grading assay. Clinical validity - A recent prospective observational study evaluated the prognostic value of both WHO 1973 and 2004 grading systems, markers CK20, FGFR3, and Ki-67, and molecular grade (combination of FGFR3 and Ki-67) in 221 patients with urothelial carcinoma. In univariate analysis, WHO

grade 1973, WHO grade 2004, pathological stage, FGFR3, Ki-67 status, and molecular grade were significantly associated with progression in stage; however, in a multivariate model, only WHO grade 1973 and 2004 remained significantly associated with progression in stage. None of the variables measured were significantly associated with recurrence-free survival (Burger 2008). Another study that included 255 patients with primary urothelial carcinoma also found that the combination of FGFR3 and Ki-67 status was not an independent predictor of recurrence-free or disease-specific survival (van Oers 2007). However, an observational study that included 286 patients with urothelial carcinoma found that in a multivariate analysis, the combination of FGFR3 and Ki-67 status predicted progression, recurrence rate, and disease-specific survival (van Rhijn 2003). Clinical utility - No studies were identified that addressed the clinical utility of the CertNDx molecular grading assay.

Conclusion: Analytic validity: No studies were identified that addressed the analytic validity of the CertNDx molecular grading assay. Clinical validity: Results from observational studies regarding the prognostic value of molecular grade (FGFR3/Ki-67) are mixed. Clinical utility: No studies were identified that addressed the clinical utility of the CertNDx molecular grading assay.

Articles: No studies were identified that addressed the analytic validity or clinical utility of the CertNDx molecular grading assay. Several studies were identified that evaluated the clinical validity of the CertNDx molecular grading assay. The most recent study was selected for review. The following study was critically appraised: Burger M, van der Aa MN, van Oers JM, et al. Prediction of progression of non-muscle-invasive bladder cancer by WHO 1973 and 2004 grading and by FGFR3 mutation status: a prospective study. *Eur Urol.* 2008;54:835-843. See [Evidence Table](#).

The use of FGFR3 for urothelial carcinoma does not meet the *Kaiser Permanente Medical Technology Assessment Criteria*.

MammaPrint Test

BACKGROUND

Breast cancer affects almost 10% of women in western countries and is a major cause of morbidity and mortality. Most patients with lymph node negative disease may be successfully treated with surgery and local irradiation. Those with more aggressive disease may benefit from adjuvant chemotherapy and hormone therapy which could significantly improve their overall and disease-free survival. It is generally accepted that breast cancer patients with the poorer prognosis would gain the most benefits from systemic adjuvant therapy. The use of this adjuvant therapy is thus one of the most critical treatment decisions during the clinical management of breast cancer patients. Currently those with aggressive breast cancer are identified according to a combination of criteria including age, clinical stage and size of the tumor, histological type and grade of cancer, axillary node status, and hormone-receptor status. The ability of these criteria to predict outcome and disease progression is imperfect. Within a given patient population at a specific predicted risk of recurrence, there are some patients whose actual clinical outcome does not match that predicted by the indicators. As a result some of those who need adjuvant therapy do not receive it, while others may receive unnecessary toxic therapy (Kallioniemi 2002, DeVigier 2002). To overcome these issues, scientists are attempting to identify more accurate prognostic indicators. Microarray technology is revolutionizing researchers' understanding of cancer biology through the simultaneous study of the expression of tens of thousands of genes. Molecular profiling is the classification of tissue or other specimens for diagnostic, prognostic, and predictive purposes based on multiple gene expression. The potential value of gene expression profiling in assessing the risk of post-surgical breast cancer recurrence has been extensively investigated over the last few years. This has led to important insights in the molecular heterogeneity of cancers by revealing biologically and clinically relevant subtypes of tumors previously indistinguishable by the conventional approaches (Bertucci 2005). Due to the biological heterogeneity of breast cancers, women with the same stage of the disease may vary widely in their response to treatment and prognosis. Several gene expression-based predictors for breast cancer have been developed, but have not been used in routine clinical practice. According to researchers, this is mainly due to the limited validation and the limited clinical description of the molecular subtypes. Validation is a major challenge for microarray studies especially those with clinical implications as it requires a large sample size and because the results are influenced by the patient selection and by choice of the methods used to analyze gene expression data (Calza 2006, Hu 2006, Ioannidis 2007). The Amsterdam 70-gene profile (MammaPrint®) was first developed using supervised gene expression profiling analysis of frozen tumor samples from two distinct patient populations. All were <55 years of age, and had lymph node negative disease. 44% had distant metastases within 5 years of completing treatment and 56% did not. By comparing the gene expression profile of patients with or without metastases, a signature 70-gene set that correlated with the outcome was identified and internally validated with the same group (van't Veer 2002), and externally validated in two retrospective groups (Van De Vijver 2002 and Buyse 2006, see evidence tables). MammaPrint® from Agendia is a qualitative in vitro diagnostic test service performed in a single laboratory using the gene expression profile of breast cancer tissue samples to assess a patient's risk for distant metastases. The MammaPrint assay uses a panel of the Amsterdam 70-gene profile described above. It is a microarray based gene expression analysis of

RNA extracted from breast tumor tissue. The MammaPrint® analysis is designed to determine the activity of specific genes in a tissue sample compared to a reference standard. Its index ranges from -1.0 to +1.0. Tumor samples with an index above the threshold of +0.4 are classified as low risk, and those with an index equal to or less than the threshold are classified as high risk. The test requires fresh frozen samples which are shipped to the Agendia reference laboratory in the Netherlands. It is performed for breast cancer patients <61 years old, with Stage I invasive breast cancer or Stage II node negative invasive breast cancer, with tumor size <5 cm. It is indicated for use by physicians as a prognostic marker only, along with other clinicopathological factors. It is not intended for diagnosis, or for predicting or detecting response to therapy, or to help select the optimal therapy for patients (FDA).

08/06/2007: MTAC REVIEW

MammaPrint Test

Evidence Conclusion: The identification and validation of gene expression panels to improve risk prediction or treatment outcomes is a multistep process that starts by 1. Identifying the candidate genes (analytic validity), followed by 2. Evaluating the genetic panel associations with risk prediction or treatment outcomes in preliminary performance studies in relevant population (clinical validity), and 3. Determining whether the use of the multigenetic assay would direct the management of patients and improve outcomes (clinical utility). The most reliable method for validation is to derive a prognostic/predictive gene set from a training set and then apply it to a completely independent set, the test set, (Simon 2003, Ionidis 2006, and Hu 2006). The MammaPrint test was developed based on research performed in the Netherlands Cancer Institute, The training set was derived from a study by van't Veer and colleagues that included 98 women < 55 years of age at diagnosis, with primary breast cancer (34 developed distant metastases within 5 years, 44 were disease free after at least 5 years). All patients were lymph node negative. 5 µg total RNA was isolated from frozen tumor material for each patient. The authors used inkjet-synthesized oligonucleotide microarrays that included 25,000 genes. Following several techniques 5000 genes were selected from the microarray, and then optimized to 70 genes with which a prognosis profile was established. The authors conducted a cross validation and concluded that a classification system based on these 70 genes outperformed all clinical variables in predicting the likelihood of distant metastases within five years. They noted however, that a selection of the patients based on the outcome (distant metastases or disease free in 5 years) was a limitation to the study. The same research team followed the initial study with a validation study (Van De Vijver, 2002) that included 295 women with either lymph node negative or lymph node positive breast cancer. The authors calculated the correlation coefficient of the level of expression of the 70- predictor genes identified in their initial study. They then classified the women with a correlation coefficient > 0.4 as having a good prognosis gene expression signature, and all the others as having a poor prognosis gene expression signature. In this validation set however the authors included 61 patients from the original training group used to derive the RNA expression signature, which could overestimate the relative risk and inflate the discriminating power of the test. The validation study included women < 55 years of age, with small tumors and at stage I or II of the disease which may not represent the entire spectrum of patients with breast cancer. Adjuvant hormone therapy or chemotherapy or both were given to most of the patients with lymph node positive disease. The Translational Research Network of the Breast International group (TRANSBIG) also conducted an independent validation study of the prognostic signatures in a retrospective series of 302 untreated patients in five European countries. The study included only women node negative early stage breast cancer who had not received systemic adjuvant therapy, and thus may not represent the all patients with breast cancer. Its overall results showed that the 70-gene signature provided prognostic information on time to distant metastases and overall survival independent of the other clinical predictors. In conclusion, the selection of the 70- predictor genes was based on analyses of tumors from patients < 55 years of age with lymph node negative cancer who do not represent all women with breast cancer. The test proved to perform well as an independent prediction tool among the selected women studied. This however, does not necessarily indicate that it would predict treatment response. To date there are no published studies that show if modification of adjuvant therapy based on this test would improve disease free or overall survival. A large randomized controlled trial (Microarray for Node negative Disease may Avoid Chemotherapy [MINDACT]) that will evaluate the clinical utility of MammaPrint is underway. The trial will directly compare the use of prognostic information provided by the standard clinicopathological criteria vs. the MammaPrint test to decide whether to offer adjuvant chemotherapy to node-negative breast cancer patients. The MINDACT plans to prospectively include 6000 women, and follow-them up for a long duration in order to determine 5-year disease free-survival rate.

Articles: The literature search revealed multiple articles on molecular and gene-expression profiling in general. For the MammaPrint test in particular, there was a published study on the training set (to develop or derive the predictive classifier or model) by Van't Veer and colleagues, and three validation studies to evaluate the predictive accuracy of the model (Van De Vijver 2002, Buyse 2006, and Glas 2006). All studies were reviewed but only the first two validation studies were critically appraised, Glas, et al's study was not selected for critical appraisal due to patient overlap with the van De Vijver study. It is to be noted that Van De Vijver, van't Veer, and several other principal

authors are named inventors on a patent application for the 70-gene signature used in the studies. All studies also had financial ties to the manufacturer. *The following studies were critically appraised:*

Van De Vijver MJ, He YD, van't Veer LJ, et al. A gene expression signature as a predictor of survival in breast cancer. *N Engl J Med* 2002;347:1999-2009. See [Evidence Table](#). Buyse M, van't Veer, L, Viale G et al on behalf of the TRANSBIG Consortium. Validation and clinical utility of a 70-gene prognostic signature for women with node negative breast cancer. *J Natl Cancer Inst* 2006;98:1183-1192. See [Evidence Table](#).

The use of the MammaPrint test in the treatment of recurring cancer does not meet the *Kaiser Permanente Medical Technology Assessment Criteria*.

OVA1™ Test for the Assessment of Suspected Ovarian Cancer

BACKGROUND

In the United States, ovarian cancer is the fifth leading cause of all cancer-related death among women. It is estimated that in 2010, there were 21,880 new cases of ovarian cancer and 13,850 deaths from ovarian cancer (Jemal 2010). The incidence of ovarian cancer increases with age with approximately two thirds of cases being diagnosed in women over the age of 55. Women with a family history of ovarian or breast cancer or who are carriers of the BRCA gene mutations are also at increased risk for ovarian cancer (Clarke-Pearson 2009). For patients with early stage disease, survival rates are greater than 90%; however, they are less than 30% for patients with advanced disease. Because of the lack of specific symptoms during the early stage approximately 70% of cases are diagnosed with advanced disease (Carter 2011). The most commonly used tests for the detection of ovarian cancer are transvaginal ultrasound (TVS) and serum CA-125. Recently, the FDA approved the OVA1™ test (Quest Diagnostics, Inc.) to be used as an adjunct to clinical/radiological evaluations for women planning surgery for an adnexal mass. This test measures the serum levels of 5 potential biochemical markers for ovarian cancer (transferrin, apolipoprotein A1, transferrin, CA-125, and β 2-microglobulin). The results of the test are then interpreted using a proprietary algorithm to yield a single score ranging from 0 to 10 to indicate the likelihood that the adnexal mass is benign or malignant. A high probability for malignancy is defined as a score of at least 5.0 in premenopausal women or 4.4 in postmenopausal women. The goal of the OVA1™ test is to provide additional information to aid in identifying patients who should be referred to a gynecologic oncologist for surgery (Carter 2011, Muller 2010). Studies suggest that women who receive their initial surgical care from an experienced gynecologic oncologist have improved outcomes and greater overall survival. Because of this the National Comprehensive Cancer Network (NCCN) recommends that all patients should undergo surgery by an experienced gynecologic oncologist (NCCN 2011). It is important to emphasize that this test is not approved for ovarian cancer screening and is not intended for use as a stand alone test. Another limitation of this test is that assay interference may occur in patients with rheumatoid factor levels of at least 250 IU/mL and triglyceride levels greater than 4.5 g/L (Muller 2010). In 2009, the FDA approved the use of this test for women over the age of 18 with an ovarian adnexal mass for which surgery is planned and have not yet been referred to an oncologist.

10/17/2011: MTAC REVIEW

OVA1™ Test for the Assessment of Suspected Ovarian Cancer

Evidence Conclusion: *Analytic validity* - No studies were identified that evaluated the analytic validity of the OVA1™ test. *Clinical validity* - A recent observational study that included 524 women with ovarian tumors who were planning to undergo surgery compared the sensitivity and specificity of physician assessment alone (which included the use of CA 125) or combined with the OVA1™ test for identifying ovarian tumors at high risk for malignancy. Results from this study suggest that the addition of the OVA1™ test to physician assessment increased the sensitivity and the negative predictive value, but decreased the specificity and the positive predictive value (Ueland 2011). The performance of the American College of Obstetricians and Gynecologists (ACOG) referral guideline for women with a pelvic mass and the effect of replacing CA 125 with the OVA1™ test was also evaluated in this study population. The substitution of CA 125 for the OVA1™ test increased the sensitivity and the negative predictive value of the ACOG guidelines, but decreased the specificity and the positive predictive value (Miller 2011). *Clinical utility* - No studies were identified that evaluated the clinical utility of the OVA1™ test. **Conclusion:** *Analytic validity:* No studies were identified that evaluated analytic validity of the OVA1™ test. *Clinical validity:* Results from a recent observational study suggest that the when added to physician assessment or substituted for CA 125, the OVA1™ test increased the sensitivity and negative predictive value of these assessments, but decrease the specificity and positive predictive value. *Clinical utility:* No studies were identified that evaluated the clinical utility of the OVA1™ test.

Articles: No studies were identified that assessed the analytic validity or clinical utility of the OVA1™ test. Two studies were identified that addressed the clinical validity of the OVA1™ test. Both of these studies were selected for review. The following studies were selected for critical appraisal: Ueland FR, Desimone CP, Seamon LG, et al. Effectiveness of a multivariate index assay in the preoperative assessment of ovarian tumors. *Obstet Gynecol*

2011; 117:1289-1297. See [Evidence Table](#). Ware Miller R, Smith A, DeSimone CP, et al. Performance of the American College of Obstetricians and Gynecologists' ovarian tumor referral guidelines with a multivariate index assay. *Obstet Gynecol*. 2011;117:1298-1306. See [Evidence Table](#).

The use of OVA1 for ovarian tumors does not meet the *Kaiser Permanente Medical Technology Assessment Criteria*.

Thyroid Nodule Gene Expression Testing (Afirma)

BACKGROUND

Thyroid nodules are very common; they are clinically identified in 5-7% of the population, and incidentally on ultrasonography in up to 50% of women and 20% of men over the age of 50. The thyroid nodules are typically benign, but 5-15% prove to be malignant. It is thus recommended that any identified nodule measuring one centimeter or more in diameter be diagnostically evaluated. Thyroid fine needle aspiration (FNA) biopsy is the most widely used method for clinical evaluation of a suspicious thyroid nodule. FNA is a safe and simple outpatient procedure that yields cellular material suitable for cytological analysis. However 15-30% of the biopsied nodules has indeterminate cytology and cannot be conclusively diagnosed by FNA biopsy alone. Most patients with indeterminate lesions (defined in the Bethesda System as Atypia of Undetermined Significance or Follicular Lesion of Undetermined Significance, suspicious for Follicular or Hurthle Cell neoplasm and suspicious for malignancy) are referred to surgery. Currently, surgery is performed for both diagnostic and therapeutic purposes in these patients with indeterminate aspirates. Surgery has high operative efficacy in removal of thyroid cancer, however approximately three-quarters of the nodules with indeterminate FNA cytology are ultimately found to be benign on final surgical pathology. Thus, a large proportion of patients with indeterminate nodules may undergo unnecessary partial or complete thyroidectomy with its potential surgical complications and risk of long-term morbidity (Alexander 2012, Duick 2012, Walsh 2012, Ali 2013). In an attempt to preoperatively classify the indeterminate thyroid nodules different novel diagnostic tests and molecular markers have been investigated. These include immunohistochemistry, mutation and gene rearrangement testing, and gene expression and microarray analysis. Each has its performance characteristics and diagnostic values. Ideally a molecular marker or panel of markers would be accurate in differentiating benign from malignant in any lesion that is considered suspicious or indeterminate. It should be simple to use, reproducible by all institutions, and cost-effective. Genetic markers associated with malignancy such as mutation markers (e.g. BRAF, RAS) and gene rearrangements (e.g. RET/PTC and PAX8-PPAry) have high specificity and positive predictive values; and when detected they can "rule in" the diagnosis of thyroid cancer. However, they have limited sensitivity and negative predictive values as they fail to detect a large proportion of malignant samples that do not contain one of the mutations or rearrangements being tested, i.e. mutation or rearrangement markers cannot 'rule out' malignancy when not detected (Alexander 2012, Kouniavsky 2012, Ward 2013). Microarray techniques seek to identify patterns of expressed RNA in the human genome that are predictive of benign or malignant thyroid disease. Unlike single gene mutations or rearrangements, microarray diagnostic tests involve tens to hundreds of expressed genes. The currently available diagnostic microarray for use in thyroid nodule analysis is the Afirma Gene Expression Classifier (GEC) recently developed by Veracyte, Inc. It is a genomic test designed with the intention of preoperative identification of benign thyroid nodules in patients with indeterminate FNA cytopathological results. The test assesses gene expression from mRNA isolated from thyroid FNA samples by comparing the mRNA expression detected in a thyroid FNA against a panel of 167 molecular genes. It uses a multidimensional algorithm to identify the thyroid FNA samples with a benign gene expression pattern (Alexander 2012, Kim 2012, Ward 2013). Afirma GEC is commercially owned by Veracyte Corporation; South San Francisco, California and is offered through a sole source, Clinical Laboratory Improvement Amendments (CLIA), a certified reference laboratory. Afirma CEC analysis is indicated only for nodules with indeterminate cytology, and is not performed on cytologically benign, malignant, or nondiagnostic (insufficient FNA samples) nodules. The assay classifies nodule as either benign or suspicious for malignancy. With a preoperative identification of a nodule that is benign rather than malignant, observation or ultrasound follow-up could be recommended instead of thyroid surgery, i.e. potentially avoids unnecessary surgery (Alexander 2012, Duick 2012, Ward 2013).

10/21/2013: MTAC REVIEW

Thyroid Nodule Gene Expression Testing (Afirma)

Evidence Conclusion: Analytic validity Evaluating the analytic performance of Afirma GEC includes studying the stability of RNA in FNAs during collection, storage, and shipment; reproducibility of the test; and its analytic sensitivity and specificity under various conditions e.g. interference of the assay with bloody FNA and genomic DNA. The literature search revealed one study (Walsh and colleagues, 2012) that evaluated the analytic performance of Afirma GEC in a number of substudies. The investigators obtained prospective FNA samples aspirated in vivo from 43 patients from outpatient clinics, preoperatively, or immediately after surgical excision.

The samples were placed in FNAProtect preservative solution and shipped chilled or frozen, then stored at -80°C upon receipt. The RNA was extracted and its yield examined for quantity and quality using positive (tissue lysate) and negative (water) as controls. Three different lots of controls were tested over several weeks of independent runs by 3 different operators to determine reliability of the test. Multiple lots of benign and malignant total RNA were manufactured and used as process controls to determine the analytic sensitivity of the test using different RNA input quantity and under different dilution of malignant FNA content. These studies indicated tolerance to variation in RNA input across a range of 5-25 ng, as well as dilution of malignant FNA material down to 20%. Analytic specificity of the test using malignant samples mixed with blood (up to 83%) and genomic DNA (up to 30%) showed minimal assay interference. However benign FNA samples mixed with relatively high proportions of blood had a potential for yielding false positive results. The authors also examined the stability of RNA in FNAs during collection and shipment and found that RNA content within FNAs preserved in FNAProtect was stable for up to 6 hours at room temperature with no change in RNA yield, and that the FNA storage and shipping temperatures had no significant effect on GEC scores. They also examined the reproducibility of the test and indicated that it was reproducible from extraction through GEC results, including variation across operators, runs, reagent lots, and laboratories. The authors concluded that the analytical performance and reproducibility of the Afirma Gene Expression Classifier was successfully verified. The research was supported by Veracyte Corporation, (the maker of Afirma GEC), and the authors of the study were either employed by, or were consultants to the corporation.

Clinical validity A perfect test would have high sensitivity and high specificity in correctly detecting or excluding a condition. A molecular diagnostic test with high sensitivity offers a high negative predictive value when the risk of malignancy (ROM) is low, and can “rule out” malignancy. Conversely, a test with high specificity offers high positive predictive value and can “rule in” cancer. To be of use in avoiding surgery, a test that better distinguishes benign from malignant nodules needs to have high sensitivity and high negative predictive value. The literature search identified two published studies on the validation of Afirma GEC (Chudova et al, 2010, and Alexander et al. 2012); both funded by Veracyte Corporation the maker of Afirma GEC. The more recent and larger validation study by Alexander and colleagues (evidence table 1), was a double-blind prospective multicenter validation study. 4,812 thyroid FNAs were obtained from 3,789 patients. 577 (12%) samples were classified as indeterminate, and less than half (46%) were ultimately selected for GEC analysis. Molecular results were compared to the gold standard of post-surgical histopathology interpreted by a panel of blinded endocrine histopathologists for clinical validation. The overall sensitivity of the Afirma test was 92% with a negative predictive value (NPV) of 93% (95% for atypical or follicular lesions of undetermined significance (AUS/FLUS), 94% for a follicular neoplasm, and 85% for a lesion suspicious for malignancy). It is to be noted that the predictive values of a test vary with the prevalence of the disease in the population studied and may not be generalized to other groups. A better analysis would be the likelihood ratios which are not affected by prevalence. Seven of the 85 (8.2%) overall cancers were diagnosed incorrectly by the GEC as benign (false negative). The authors attributed the false negative results to insufficient RNA in the FNA sample used for GEC. The test had an overall low specificity and positive predictive values (52% and 47% respectively). Atypical or follicular lesions of undetermined significance (AUS/FLUS) accounted for almost 50% of the indeterminate thyroid FNAs samples. 43% of these FNA were reclassified with the GEC as benign and 57% remained in their suspicious category. Other investigators showed that repeat FNAs without a molecular test can also accurately reclassify >50% of the nodules in the AUS/FLUS category as benign (Faquin 2013). The study was conducted to validate the GEC accuracy by comparing it to surgical histopathology, and the authors did not compare its performance to repeat FNA or other immunochemical testing.

Clinical utility The clinical utility of Afirma GEC was evaluated in a retrospective study by Duick and colleagues, 2012, (Evidence table 2). They obtained their data from 21 endocrinology practices in 11 states. The authors conducted a chart review of 368 patients with 395 cytologically indeterminate thyroid nodules that were GEC benign. 7.6% of these patients with Afirma GEC benign nodules underwent surgery and 94.4% were managed nonoperatively. The study did not have a comparison group, but the authors compared the 7.6% surgical rate to a 74% historical rate of diagnostic surgery (P<0.001). The indications for surgery for those with GEC benign results included a large size or rapid growth of the nodules, local pressure symptoms, or the presence of a second suspicious nodule or malignant nodule. The authors explained that these were similar to indications for surgery on nodules with benign FNA cytologically. The study was retrospective, used a historical comparison, and investigated the decision-making of endocrinologists experienced in managing patients with thyroid nodules, which may differ from that made by primary care providers or other specialists. In addition the authors of the study did not provide data on long-term follow-up of those who were managed by watchful waiting rather than surgery. In conclusion, there is insufficient evidence to determine whether Afirma GEC is more accurate than repeat FNA or immunochemical testing in reclassifying cytologically indeterminate thyroid nodules. There is also insufficient evidence to determine the impact of Afirma GEC on clinical management and net health outcomes in patients with indeterminate thyroid nodules.

Articles: The literature search for gene expression classifier for preoperative identification of benign thyroid nodules with indeterminate fine needle aspiration cytopathology revealed a number of articles on molecular diagnostic tests. Many were reviews, editorials, letters, or were unrelated to the current review. The search identified a study on the analytic validity of the test, two on its clinical validity, and retrospective study on its clinical

utility. The following studies were selected for critical appraisal. Alexander EK, Kennedy GC, Baloch ZW, et al. Preoperative diagnosis of benign thyroid nodules with indeterminate cytology. *N Engl J Med.* 2012;367:705-715. [See Evidence Table](#) Duick DS, Klopper JP, Diggans JC, et al. The impact of benign gene expression classifier test results on the endocrinologist-patient decision to operate on patients with thyroid nodules with indeterminate fine-needle aspiration cytopathology. *Thyroid.* 2012 22:996-1001. [See Evidence Table](#)

The use of does Afirma® Thyroid FNA Analysis (Gene Expression Classifier) for Thyroid Nodules with Indeterminate Fine Needle Aspiration Cytopathology not meet the *Kaiser Permanente Medical Technology Assessment Criteria*.

DecisionDx - Melanoma

BACKGROUND

Skin cancer is extremely common accounting for nearly half of all cancers in the United States. Melanoma, the most aggressive type of skin cancer, occurs as a result of abnormal melanocytes, most often caused by over- exposure to ultraviolet radiation from the sun. When detected early, cutaneous melanoma can be surgically excised resulting in a 5-year overall survival rate of 91%-97%. Despite these odds, however, the clinical behavior of cutaneous melanoma is highly variable and some melanomas, that appear less risky, will develop into advanced disease and require extensive treatments such as additional surgery, immunotherapy, targeted therapy, chemotherapy and radiation therapy (ACS 2015). As with all cancers, a primary challenge is predicting prognosis. Conventional methods of melanoma staging are characterized by the American Joint Committee on Cancer (AJCC) TNM System. The TNM system specifically refers to Tumor thickness, spread to nearby lymph Nodes, and Metastasis. Based on history and physical exam, as well as, biopsy, imaging and pathology, the TNM system groups patients with melanoma into stages, 0-IV based on the advanced nature of the disease (Balch, Gershenwald et al. 2009). The stage of the melanoma is an estimate of prognosis and will ultimately guide treatment options. Recently, gene expression profiling (GEP) has been proposed for use in cancer management. The technique specifically analyzes the patterns of genetic material contained in tumor cells and has the potential ability to predict clinical outcomes associated with cancer. One such test, the DecisionDx-Melanoma™, developed by Castle Biosciences Inc. (Friendswood, TX), is described to more accurately classify stage I and II melanoma.

Proposed as an adjunct to conventional staging systems, the DecisionDX-Melanoma test includes 31 genes, 28 of which have previously been associated with melanoma and the remaining three, controls (Winnepeninckx, Lazar et al. 2006). The results of the DecisionDx-Melanoma test is further claimed to stratify stage I and II melanomas into one of two classes; class one identifying patients as low risk of metastasis, or class two indicating high risk. The developer claims that the information provided by the DecisionDx-Melanoma test enables physicians to tailor, patient specific, surveillance and treatment plans informing, for example, the intensity of surveillance, need for referral to specialists, evaluation of adjuvant treatments and clinical trial eligibility (CastleBiosciencesInc. 2015).

04/20/2015: MTAC REVIEW

DecisionDx - Melanoma

Evidence Conclusion: The study aimed to develop a prognostic genetic signature based on previous analyses of cutaneous melanoma tumors. To do this, the investigators included 268 archived tissue samples and divided the sample into two cohorts, development (n=164) or validation (n=104). The investigators compared the patient clinical outcomes at five years with the GEP test prediction. Overall, Kaplan-Meier analysis indicated that the five year disease free survival (DFS) rates in the validation cohort were 97% and 31% for predicted class 1 and 2, respectively (p<0.0001). These results were comparable to the DFS rated in the development cohort, 100% and 38% for class 1 and class 2, respectively (p<0.0001). The investigators ultimately concluded that in patients with primary cutaneous melanoma, the GEP signature accurately predicts metastasis risk (Gerami, Cook et al. 2015). [Evidence Table 1] The investigators had the clear intent to develop and validate a GEP for predicting metastatic risk in stage I and II cutaneous melanoma. The patient sample was well defined and the study design, cohort, appeared to be appropriate for the development of the genetic signature. To validate the test, however, the study relied on archived tumor samples with at least five years of follow-up. While this is a sufficiently long time to detect the outcome of interest, and the investigators used an independent sample, a prospective study would be a more appropriate design for validation. With that said, the investigators report that samples were collected at a similar point in the course of the disease, diagnosis, however the diseases progression at diagnosis may have varied between patients and it is not clear if the investigators were blinded to prognostic factors. On a final note, the study was funded by the test manufacturer and at least two of the investigators have financial ties with Castle Biosciences, Inc. Conclusions: There is limited evidence to conclude that the DecisionDx-Melanoma test is valid. There is insufficient evidence to conclude that the DecisionDx-Melanoma test has prognostic accuracy in predicting metastatic risk. There is insufficient evidence to conclude that the DecisionDx-Melanoma test is not harmful to patients. There is insufficient evidence to establish the clinical utility and therapeutic impact of the DecisionDx-Melanoma test.

Articles: The literature search was carried out to identify studies relating to the prognostic value of the DecisionDx-Melanoma test. The search revealed a variety of publications discussing the use of GEP and one publication identifying the genes associated with melanoma progression and prognosis (Winnepeninckx, Lazar et al. 2006). No studies were identified in which the DecisionDX-Melanoma was prospectively analyzed and followed- up in populations with Stage I and II melanoma. A search of the NIH Clinical Trials database identified two manufacturer sponsored prospective studies currently in the enrollment stage. The best, currently available, evidence was a development and validation study published by Castle Biosciences, Inc. The following articles were selected for critical appraisal: Gerami P, Cook RW, Wilkinson J, et al. Development of a prognostic genetic signature to predict the metastatic risk associated with cutaneous melanoma. *Clinical Cancer Research*. 2015;21(1);175-183. See [Evidence Table](#).

The use of DecisionDx-Melanoma does not meet the *Kaiser Permanente Medical Technology Assessment Criteria*.

OVA1 Assessment for Ovarian Cancer

BACKGROUND

Ovarian cancer is the most lethal gynecological malignant worldwide. The five-year overall survival is over 90% in patients with stage I disease and only 20-40% for stages III and IV. Unfortunately, because of the lack of specific symptoms during the early stage approximately 70% of cases present with an advanced stage disease. Detection of ovarian cancer at an early stage would have a significant impact on reducing mortality, however to date; there is no screening or biomarker test that meets the criteria for a beneficial screening test in asymptomatic women with early ovarian cancer (Carter 2011, Cohen 2014, Leung 2014). Serum CA-125, a high molecular weight glycoprotein, remains the most widely used biomarker for the confirmation of diagnosis and management of ovarian cancer. Serum CA-125 however, is more prominently expressed in patients with late stage serous tumors; it is elevated in 50-60% of women with stage I epithelial ovarian cancer, and in 75-90% of patients with advanced stage disease. Elevated circulating CA-125 has also been documented in uterine fibroids, endometriosis, pregnancy, menstruation, benign ovarian neoplasms, liver cirrhosis, and other malignancies making it a less useful marker for the detection of ovarian cancer (Autelitano 2012, Cohen 2014). Improvements have been made in the preoperative diagnosis of ovarian cancer by combining serum CA-125 concentration with ultrasound score and menopausal status, into a Risk of Malignancy Index (RMI) which was found to outperform CA-125 alone in discriminating between a benign and malignant pelvic mass. Over the past two decades diagnostic triage methods incorporating clinical algorithms, serum biomarkers, imaging, or a combination of these techniques have been investigated to improve its diagnostic efficiency in predicting ovarian malignancy in women with adnexal masses. The Risk of malignancy Algorithm (ROMA) and OVA1 test are two algorithms recently developed for the assessment of malignancy risk in these women. These are not screening tests, but are potential tools to further triage women to the appropriate provider once the decision for surgical intervention has been made (Autelitano 2012, Bristow 2013, Cohen 2014). Combining multiple variables or markers in a single biomarker assay (in vitro diagnostic multivariate assay [IVDMIA, or MIA]) has the potential advantage of complementing the information provided by a single-valued index. The inclusion of biomarkers in an IVDMIA requires that they are complementary and collectively outperform a single marker with respect to its intended uses. CA-125 remains the best tumor marker, and the selection of additional biomarkers is based mainly on their ability to detect malignancy in cancer patients with low CA-125 level or to reduce false positive results among non-cancer patients with elevated serum CA-125 levels (Zhang 2012). Ova1™ test (developed by Vermillion and licensed to Quest Diagnostics, Inc.) is the first IVDMIA of protein biomarkers cleared by the FDA to be used as an adjunct to clinical and radiological evaluations for women over the age of 18 who have planned to undergo surgery for an adnexal mass and have not been referred to a gynecologic oncologist. Studies suggest that women who receive their initial surgical care from an experienced gynecologic oncologist are more likely to have better outcomes including surgical staging, optimal debulking, and improved median and overall-5-year survival. Ova1™ test is a qualitative test that measures the serum levels of 5 potential biochemical markers for ovarian cancer (CA-125, prealbumin, apolipoprotein A-1, β 2-microglobulin, and transferrin). The results of the test are then interpreted using a proprietary algorithm to yield a single score ranging from 0 to 10 to indicate the likelihood that the adnexal mass is benign or malignant. A high probability for malignancy is defined as a score of ≥ 5.0 in premenopausal women or ≥ 4.4 in postmenopausal women. The decision for selecting these cutoff values was made to emphasize the need for high sensitivity to minimize the risk of false negative results for patients who actually have a malignant lesion. A limitation to OVA1™ is that all the included markers with the exception of CA-125 are acute phase reactants that may be nonspecific for ovarian cancer. Another limitation is interference of triglyceride levels greater than 4.5g/L or rheumatoid factor levels more than 250IU/mL with the biomarkers assay (Muller 2010, Carter 2011, Zheng 2012, Leung 2014).

04/20/2015: MTAC REVIEW

OVA1 Assessment for Ovarian Cancer

Evidence Conclusion: The main purpose of adding biomarkers to an established tumors biomarker as CA-125, in a multivariate index assay (MIA), is to achieve a very high sensitivity without sacrificing the specificity. However, the published studies evaluating OVA1™ showed the test improved the sensitivity of the physicians' assessment in predicting ovarian malignancy in women with adnexal masses, but at the cost of reducing the specificity and positive predictive value. The FDA cleared the OVA1™ test based on the results of Ueland and colleagues' study that was reviewed earlier by MTAC in 2011. The study compared the sensitivity, specificity, and predictive values of physician assessment with or without adding the multivariate index assay (MIA) in identifying high-risk ovarian tumors. The study enrolled 590 women (524 evaluable with both MIA and CA-125-II) with a documented ovarian mass on imaging and planned surgery within 3 months of imaging. 53% of the women were enrolled by non-gynecologic oncologists and the rest by gynecological oncologist. The MIA index assay test was performed on preoperative serum samples, and the results were correlated with preoperative physician assessment. There was no specific protocol for the clinical assessment. Using surgical pathology as the gold standard, 161 women were diagnosed with a malignant and 363 with a benign ovarian tumor. The results of the analysis showed that the sensitivity of non-gynecologic oncologists' assessment increased from 72% to 92% with the addition of the MIA test (78% and 99% respectively for gynecologic oncologists). The negative predictive value increased slightly with the addition of the MIA test. On the other hand, the specificity and positive predictive values dropped significantly with the addition of the assay (the specificity was reduced from 83% to 42% for non-gynecologic oncologists and from 75% to 26% for gynecologic oncologists and the positive predictive value dropped from 60% to 36% and from 63% to 43% in the two groups of respectively). The studies published after that pivotal study were conducted mainly by the same group of investigators who either analyzed the results of women enrolled in some or all 44 sites participating in the study. The studies were sponsored by Vermillion Inc., and the investigators had financial ties to the company. The largest and most recent of these studies (Longoria et al 2014) (Evidence table 1) compared the accuracy and predictive values of the multivariate index assay, OVA1™ to clinical assessment, CA-125-II, and the modified American Congress of Obstetricians and Gynecologists (ACOG) guidelines, for the detection of early-stage ovarian cancer in 1,016 women undergoing surgery for an adnexal mass. The authors did not indicate whether the assessors were blinded to the other tests and/or clinical evaluation results. The study did not include women without adnexal masses or with other disorders that may lead to elevated levels of CA-125 or any of the other biomarkers included in the assay. Overall, similar to the Ueland and colleagues' study, as well as the other published studies using MIA test, Longoria, et al's study showed that the addition of OVA1™ to clinical assessment may significantly improve the sensitivity of detecting early-stage ovarian cancer, but at the expense of reducing the specificity, which would result in referral of more patients with benign conditions to gynecologic oncologists for surgery. The overall results of the study show the following: **Comparative performance for evaluable women in all cancer cases (from evidence table 1)**

	Sensitivity %	Specificity %	PPV %	NPV %
OVA1	92.2%	49.4%	37.9%	94.9%
Clinical assessment*	74.5%	86.3%	64.6%	91.0%
OVA1 + clinical assessment	95.3%	44.2%	36.4%	96.6%
CA 125-II	70.6%	89.6%	69.5%	90.1%
Modified ACOG guidelines**	80.0%	76.5%	53.3%	91.9%

*The authors did not clearly explain that clinical assessment included CA125-II for all women

** Included: very elevated CA125 (>67U/mL), ascites, and evidence of abdominal or distant metastasis for premenopausal women.

For postmenopausal women the ACOG criteria were Elevated CA125 (>35 u/mL, nodular or fixed pelvic mass, ascites, and evidence of abdominal or distant metastasis.

The studies had enrolled selected groups of women with adnexal masses who were referred to surgery in multiple centers with no standardized process for data collection or referral practice. The referral pattern was retrospectively analyzed, and the impact of the test on health outcomes was not evaluated. In addition, the studies were funded by Vermillion Inc, the developer of the test, and the principal investigators had financial ties to the company. The performance of OVA1™ was not compared to other risk assessment algorithms as ROMA, ultrasound-based risk assessment models, or other diagnostic tools that may lead to similar sensitivity and superior specificity to OVA1™. Conclusion: The published studies do not provide sufficient evidence to determine the clinical utility and impact of using OVA1™ assay on health outcomes of women with ovarian tumors.

The use of OVA 1 does not meet the *Kaiser Permanente Medical Technology Assessment Criteria*.

Date Created	Date Reviewed	Date Last Revised
1997	10/04/2011 ^{MDCRPC} , 8/07/2012 ^{MDCRPC} , 11/06/2012 ^{MDCRPC} , 04/02/2013 ^{MDCRPC} , 05/07/2013 ^{MDCRPC} , 06/04/2013 ^{MDCRPC} , 08/06/2013 ^{MPC} , 03/04/2014 ^{MPC} , 06/03/2014 ^{MPC} , 07/01/2014 ^{MPC} , 10/07/2014 ^{MPC} , 11/04/2014 ^{MPC} , 02/03/2015 ^{MPC} , 10/04/2016 ^{MPC}	04/04/2017

^{MDCRPC} Medical Director Clinical Review and Policy Committee
^{MPC} Medical Policy Committee

Revision History	Description
05/11/2015	Array-Based Comparative Genomic Hybridization (aCGH): Removed MCG and reactivated GHC insufficient evidence criteria
06/02/2015	MPC approved MTAC recommendation of insufficient evidence for OVA1 & DecisionDx-Melanoma Testing
06/04/2015	Added Cologuard
06/30/2015	Added LCD link for cytogenetic studies
08/27/2015	Add LCD for CYP Genes
09/08/2015	Revised LCD CYP2C19 (CPT-81225), CYP2D6 (CPT-81226), CYP2C9(CPT81227), and VKORC1(CPT-81355) Genetic Testing (L36311), Cytogenetic Studies L34067
10/13/2015	Added Medicare molecular testing LCD
10/27/2015	Added codes that do not need review
11/18/2015	Added Medicare MoIDX links
03/01/2016	Discontinue review for Factor II & V
08/30/2016	Combined Risk Prognosticator Test to Genetic Screening criteria
09/06/2016	Added Prostate Cancer Gene Expression Testing- Onctotype DX MCG A-0712 to criteria
10/24/2016	Changed Veristat to match Pharmacogenomic policy
11/01/2016	MPC approved to accept the genetic testing recommendations from the MCG 20 th edition as outlined
01/23/2017	Added LCD 36544 & LCD 36186
04/04/2017	Added MTAC Review
05/16/2017	Added Percepta LCD

Codes

CPT: 81161, 81162, 81200, 81201, 81202, 81203, 81204, 81205, 81208, 81209, 81210, 81211, 81212, 81213, 81214, 81215, 81216, 81217, 81218, 81219, 81222, 81223, 81224, 81225, 81226, 81227, 81228, 81229, 81235, 81242, 81243, 81244, 81245, 81246, 81250, 81251, 81252, 81253, 81254, 81255, 81256, 81257, 81260, 81262, 81263, 81264, 81265, 81266, 81267, 81268, 81270, 81272, 81273, 81275, 81276, 81280, 81287, 81288, 81290, 81291, 81292, 81293, 81294, 81295, 81296, 81297, 21298, 81299, 81300, 81301, 81302, 81303, 81304, 81310, 81311, 81313, 81314, 81315, 81316, 81317, 81318, 81319, 81321, 81322, 81323, 81324, 81325, 81326, 81327, 81330, 81331, 81332, 81350, 81355, 81370, 81371, 81372, 81373, 81374, 81375, 81376, 81377, 81381, 81382, 81383, 81400, 81401, 81402, 81403, 81404, 81405, 81406, 81407, 81408, 81410, 81411, 81412, 81413, 81414, 81415, 81416, 81417, 81420, 81422, 81425, 81426, 81427, 81430, 81431, 81432, 81433, 81434, 81435, 81436, 81437, 81438, 81439, 81440, 81442, 81445, 81450, 81455, 81460, 81465, 81470, 81471, 81479, 81490, 81493, 81504, 81525, 81538, 81545, 81599, 84999, 0006M, 0007M, 0008M
 Veristat – 81538, 81599

These codes do not need review - 81206, 81207, 81220, 81221, 81240, 81241, 81261, 81340, 81341, 81342, 81372, 81374, 81375, 81376, 81377, 81378, 81379, 81380