



Hereditary Cancer Genetic Test

Results to date: April 28, 2016

Executive Summary

Color has developed a next-generation sequencing-based test for hereditary cancer. This test analyzes 30 genes associated with increased risk to develop breast, ovarian, colorectal, melanoma, pancreatic, prostate, stomach, and uterine cancers (Supplemental Table 1). The assay fully sequences the coding sequences and intron/exon boundaries for the genes of interest, with the exceptions noted below. Several intronic regions are also included in order to improve the resolution of copy number variation detection. The assay has a high degree of analytical validity for the detection of single nucleotide variants, small insertions and deletions (indels), and larger deletions and duplications (copy number variants, or CNVs). Validation using 507 blinded clinical specimens and 34 cell lines yielded an accuracy of 100% for 522 variants representing all these classes.

Introduction

Sequencing the first human genome took more than 10 years and \$2.7 billion dollars. However, sequencing technologies have evolved tremendously over the last decade, enabling assessment of genetic aberrations in routine clinical practice¹⁻⁶. In April 2015, Color launched a test with 19 genes in which pathogenic mutations have been associated with an elevated risk for breast and ovarian cancer. In the hereditary cancer genetic test, Color uses the same clinical-grade, quality-controlled sequencing platform to analyze the risk of developing hereditary cancer due to inheritance of a pathogenic mutation in 30 cancer predisposition genes⁷.

Materials & Methods

Color laboratory, certified by CLIA (05D2081492) and accredited by CAP (8975161), has developed a systematic process of automated laboratory protocols and tailored bioinformatics analysis to achieve reliable next-generation sequencing (NGS) results. This process is based on laboratory products from industry leaders such as Agilent, Illumina and Hamilton. Specifically, it includes target enrichment by Agilent's SureSelect method (v1.7) and sequencing by Illumina's NextSeq 500 (paired-end 150bp, High Output kit). At

several points along the process, automated quality control checks have been incorporated to ensure sample identification, high quality of DNA isolation, library preparation, target capture, and sequencing. In addition, each sequencing test contains two fully-characterized positive controls. The bioinformatics pipeline was built using well-established algorithms such as BWA-MEM, SAMtools, Picard and GATK. CNVs are detected using dedicated internally developed algorithms for read depth analysis and split-read alignment detection. Variants are classified according to the standards and guidelines for sequence variant interpretation of the American College of Medical Genetics and Genomics/ACMG⁸. Variant classification categories include pathogenic, likely pathogenic, variant of uncertain significance (VUS), likely benign, and benign. All variants are evaluated by a board certified medical geneticist or pathologist. Variants classified as pathogenic or likely pathogenic are always confirmed by a secondary technology (Sanger sequencing, aCGH or MLPA) before getting reported.

The Color hereditary cancer genetic test analyzes 30 genes in which genetic alterations have been associated with an elevated risk for breast, ovarian, colorectal, melanoma, pancreatic, prostate, uterine and stomach cancer (Supplemental Table 1). These genes are APC, ATM, BAP1, BARD1, BMPR1A, BRCA1, BRCA2, BRIP1, CDH1, CDK4, CDKN2A (*p14ARF* and *p16INK4a*), CHEK2, EPCAM, GREM1, MTF, MLH1, MSH2, MSH6, MUTYH, NBN, PALB2, PMS2, POLD1, POLE, PTEN, RAD51C, RAD51D, SMAD4, STK11, and TP53. The majority of these genes are assessed for variants within all coding exons (+/- 20bp flanking each exon). Additionally, non-canonical splice regions are also included. Several regions that cannot be reliably assessed with standard target enrichment protocols, such as exons 12-15 of PMS2 and several non-coding homopolymers, were excluded. For the CDK4, MTF, POLD1 and POLE genes, the elevated risk of cancer is associated with distinct functional genomic regions. For this reason, only the following regions are analyzed (genomic coordinates in GRCh37):
CDK4 - chr12:g.58145429-58145431 (codon 24)⁹⁻¹¹
MTF - chr3:g.70014091 (including c.952G>A)¹²⁻¹⁴
POLD1 - chr19:g.50909713 (including c.1433G>A)^{15,16}
POLE - chr12:g.133250250 (including c.1270C>G)^{15,16}

In *EPCAM*, only large deletions and duplications that include the 3' end of the gene are reported. These are the only variants known to silence the *MSH2* gene and therefore increase risk of associated cancer^{17,18}. *GREM1* is only tested for duplications in the upstream regulatory region¹⁹⁻²¹.

Our validation strategy adhered to guidelines for NGS from the College of American Pathologists (CAP), the ACMG²², the Clinical and Laboratory Standards Institute²³, the Nex-StoCT workgroup for Standardization of Clinical Testing by NGS²⁴ and FDA Standards for NGS²⁵. The validation study included saliva samples, well-characterized cell lines and DNA specimens, previously extracted from blood from patients who had been diagnosed with hereditary cancer and whose genetic variants had been previously characterized elsewhere (Table 1). Together these groups constitute a good representation of the possible variant types across the 30 genes in the Color hereditary cancer genetic test.

Study 1: Reference materials with public data

Every sequencing run contains two positive controls (NA12878 and NA19240), which have been recommended as reference materials by the National Institute of Standards and Technology/NIST²⁶. In addition, Color has sequenced the Ashkenazi Jewish father-mother-son trio NA24149, NA24143 and NA24385. Variant calls in these reference materials were compared against the union of reported variants by NIST [NCBI Get-RM] and Complete Genomics^{27,28}. Several low-confidence variants in the NIST and Complete Genomics datasets were confirmed by Sanger sequencing at an independent laboratory.

Table 1a, study 1: **Overview of variants, stratified by variant type: NIST reference materials²⁶.**

Specimen	Number of non-pathogenic variants			Total
	SNVs	Indels	CNVs	
NA12878	60	4	0	64
NA19240	55	4	0	59
NA24143	46	2	0	48
NA24149	55	3	0	58
NA24385	54	2	0	56
Total	270	15	0	285

Study 2: Blinded specimens from patients with personal history of cancer

The 30-gene hereditary cancer genetic test was validated using two groups of patients who had previously been diagnosed with cancer. The first group consisted of 29 cell lines [Coriell Institute for Medical Research and American Type Culture Collection, ATCC], many of which carry pathogenic variants in *BRCA1* and *BRCA2*. The second group consisted of 507 anonymized DNA specimens provided by Mary-Claire King, Ph.D. and Tom Walsh, Ph.D. Of these 507 specimens, 183 specimens had pathogenic variants previously identified in at least one of the 30 genes^{3-5,29-36}, and the other 324 specimens had tested negative for germline variants in the same genes. Importantly, these clinical samples were provided to Color in a “blinded” manner; i.e. Color did not have information regarding the status or genetic makeup of the samples other than the past cancer history. After the Color test was performed, results were submitted to our collaborators to be compared against the previously identified variants. This allowed Color to test the accuracy of its assay in the absence of any a priori knowledge of genetic variants.

Table 1b, study 2: **Overview of pathogenic and likely pathogenic variants, stratified by variant type: 29 cell lines [Coriell Institute and American Type Culture Collection] and 507 clinical samples.**

Specimen	Number of pathogenic variants			Number of likely pathogenic variants			Total
	SNVs	Indels	CNVs	SNVs	Indels	CNVs	
Cell lines (n=29)	14	18	NA	4	1	NA	37
Clinical samples, blinded group (n=507)	65	69	43	16	1	6	200
Total	79	87	43	20	2	6	237

Study 3: Independent confirmation of variants in consecutive Color cohort

As part of Color’s quality control system, a set of 640 variants was submitted for confirmation by Sanger sequencing. This set contains 206 variants, detected in the initial consecutive cohort of Color’s 19-gene breast and ovarian cancer genetic test, that had been classified as pathogenic or likely pathogenic.

Study 4: Technical precision: reproducibility and repeatability

Precision of the Color hereditary cancer genetic test was assessed with 3 replicate runs, which were performed by different operators. These runs used multiple lot numbers of critical reagents such as DNA polymerase and baits as well as multiple thermocyclers and sequencers. Intra-assay repeatability was computed by comparing results for 22 unique samples that had been replicated multiple times within the same run. Inter-assay reproducibility was assessed by comparing results for 61 unique samples that had been replicated multiple times across different runs. These precision measurements were calculated using all detected variants, independent of variant type (SNV/indel/CNV), classification and confirmation.

Table 2, studies 1-2: **Assessment of accuracy in detection of rare single nucleotide variants, insertions/deletions and copy number variants.**

Study	Specimen	Number of variants	True Positives	False Positive *	False Negative*
1	NA12878	64	64	0	0
	NA19240	59	59	0	0
	NA24143	48	48	0	0
	NA24149	58	58	0	0
	NA24385	56	56	0	0
2	Coriell/ ATCC cell lines (n=29)	37	37	0	0
2	Blinded samples (n=507)	200	200	0	0
Total	541	522	522	0	0

*Assessment of False Positives and False Negatives was based on all variants in the reportable range for the recommended NIST reference materials (Table 1a) and all (likely) pathogenic variants in the remaining validation specimens.

Results

The Color hereditary cancer genetic test had proven analytical validity and 100% concordance with known variants in all 30 genes across 507 previously sequenced clinical samples and 34 cell lines. The 522 variants identified in previous clinical testing, including SNVs, small indels, and CNVs, were correctly detected in a blinded analysis. In this dataset, 237 variants had been classified as pathogenic or likely pathogenic, while no false positive pathogenic variants were called in any of these 541 samples (Table 2). In addition, all 640 germline variants submitted for Sanger sequencing were confirmed and no additional variants of relevance were detected (Table 3). Repeatability within-run amounted to 100% over 1212 variants (Jeffreys 95% Confidence Interval: 0.998-1), while reproducibility between-runs was 9613 of 9615 variants (99.98%, 95% CI: 0.999-1, see Table 4).

Table 3, study 3: **Overview of secondary confirmation results by Sanger sequencing for 640 variants, of which 206 variants had been classified as likely pathogenic or pathogenic in a consecutive cohort of patients taking the Color 19-gene genetic test for breast and ovarian cancer.**

Gene	Total	True Positives	False Positives	False Negatives
ATM	94	94	0	0
BARD1	29	29	0	0
BRCA1	48	48	0	0
BRCA2	85	85	0	0
BRIP1	38	38	0	0
CDH1	23	23	0	0
CHEK2	77	77	0	0
MLH1	21	21	0	0
MSH2	52	52	0	0
MSH6	50	50	0	0
NBN	24	24	0	0
PALB2	32	32	0	0
PMS2	24	24	0	0
PTEN	2	2	0	0
RAD51C	13	13	0	0
RAD51D	11	11	0	0
STK11	8	8	0	0
TP53	9	9	0	0
Total	640	640	0	0

Table 4: Overview of Color Test performance across validation studies 1-4.

	Studies	Results	Score [Jeffreys 95% CI]
Accuracy	1-2	541/541 samples	100% [0.995-1]
Sensitivity	1-2	522/522 variants	100% [0.995-1]
Specificity	1-2	0 FPs** in 541 samples	100% [0.995-1]
PPV*	1-3	0 FPs** in 522+640=1162 variants	100% [0.998-1]
Repeatability	4	1212/1212 variants	100% [0.998-1]
Reproducibility	4	9613/9615 variants***	99.98% [0.999-1]

*PPV = Positive Predictive Value

**FP = False Positive

***Two likely benign variants, located in a homopolymer repeat and in a region of high GC content, were not reproduced in all replicates.

Conclusion

The blinded validation studies 1-2 yielded 100% accuracy [95% confidence interval 99.5% - 100%] of the Color hereditary cancer genetic test based on a set of 369 SNVs, 104 indels, and 49 CNVs. In addition, 640 variants (study 3) were confirmed independently by Sanger sequencing. Similar validation studies are ongoing to expand test results with rare and technically challenging variants.

References

- Lincoln SE, Kobayashi Y, Anderson MJ, et al. A Systematic Comparison of Traditional and Multigene Panel Testing for Hereditary Breast and Ovarian Cancer Genes in More Than 1000 Patients. *J Mol Diagn.* 2015;17(5):533-544.
- Susswein LR, Marshall ML, Nusbaum R, et al. Pathogenic and likely pathogenic variant prevalence among the first 10,000 patients referred for next-generation cancer panel testing. *Genet Med.* December 2015. doi:10.1038/gim.2015.166.
- Shirts BH, Casadei S, Jacobson AL, et al. Improving performance of multigene panels for genomic analysis of cancer predisposition. *Genet Med.* February 2016. doi:10.1038/gim.2015.212.
- Walsh T, Lee MK, Casadei S, et al. Detection of inherited mutations for breast and ovarian cancer using genomic capture and massively parallel sequencing. *Proc Natl Acad Sci U S A.* 2010;107(28):12629-12633.
- Pritchard CC, Salipante SJ, Koehler K, et al. Validation and implementation of targeted capture and sequencing for the detection of actionable mutation, copy number variation, and gene rearrangement in clinical cancer specimens. *J Mol Diagn.* 2014;16(1):56-67.
- LaDuca H, Stuenkel AJ, Dolinsky JS, et al. Utilization of multigene panels in hereditary cancer predisposition testing: analysis of more than 2,000 patients. *Genet Med.* 2014;16(11):830-837.
- Rahman N. Realizing the promise of cancer predisposition genes. *Nature.* 2014;505(7483):302-308.
- Richards S, Aziz N, Bale S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med.* 2015;17(5):405-424.
- Wölfel T, Hauer M, Schneider J, et al. A p16INK4a-insensitive CDK4 mutant targeted by cytolytic T lymphocytes in a human melanoma. *Science.* 1995;269(5228):1281-1284.
- Zuo L, Weger J, Yang Q, et al. Germline mutations in the p16INK4a binding domain of CDK4 in familial melanoma. *Nat Genet.* 1996;12(1):97-99.
- Soufir N, Avril MF, Chompret A, et al. Prevalence of p16 and CDK4 germline mutations in 48 melanoma-prone families in France. The French Familial Melanoma Study Group. *Hum Mol Genet.* 1998;7(2):209-216.
- Bertolotto C, Lesueur F, Giuliano S, et al. A SUMOylation-defective MITF germline mutation predisposes to melanoma and renal carcinoma. *Nature.* 2011;480(7375):94-98.
- Yokoyama S, Woods SL, Boyle GM, et al. A novel recurrent mutation in MITF predisposes to familial and sporadic melanoma. *Nature.* 2011;480(7375):99-103.

14. Ghiorzo P, Pastorino L, Queirolo P, et al. Prevalence of the E318K MITF germline mutation in Italian melanoma patients: associations with histological subtypes and family cancer history. *Pigment Cell Melanoma Res.* 2013;26(2):259-262.
15. Palles C, Cazier J-B, Howarth KM, et al. Germline mutations affecting the proofreading domains of POLE and POLD1 predispose to colorectal adenomas and carcinomas. *Nat Genet.* 2013;45(2):136-144.
16. Valle L, Hernández-Illán E, Bellido F, et al. New insights into POLE and POLD1 germline mutations in familial colorectal cancer and polyposis. *Hum Mol Genet.* 2014;23(13):3506-3512.
17. Ligtenberg MJL, Kuiper RP, Chan TL, et al. Heritable somatic methylation and inactivation of MSH2 in families with Lynch syndrome due to deletion of the 3' exons of TACSTD1. *Nat Genet.* 2009;41(1):112-117.
18. Kovacs ME, Papp J, Szentirmay Z, Otto S, Olah E. Deletions removing the last exon of TACSTD1 constitute a distinct class of mutations predisposing to Lynch syndrome. *Hum Mutat.* 2009;30(2):197-203.
19. Jaeger E, Leedham S, Lewis A, et al. Hereditary mixed polyposis syndrome is caused by a 40-kb upstream duplication that leads to increased and ectopic expression of the BMP antagonist GREM1. *Nat Genet.* 2012;44(6):699-703.
20. Rohlin A, Eiengård F, Lundstam U, et al. GREM1 and POLE variants in hereditary colorectal cancer syndromes. *Genes Chromosomes Cancer.* 2016;55(1):95-106.
21. Davis H, Irshad S, Bansal M, et al. Aberrant epithelial GREM1 expression initiates colonic tumorigenesis from cells outside the stem cell niche. *Nat Med.* 2015;21(1):62-70.
22. Rehm HL, Bale SJ, Bayrak-Toydemir P, et al. ACMG clinical laboratory standards for next-generation sequencing. *Genet Med.* 2013;15(9):733-747.
23. CLSI. *Nucleic Acid Sequencing Methods in Diagnostic Laboratory Medicine; Approved Guideline.* Published April 2014.
24. Gargis AS, Kalman L, Berry MW, et al. Assuring the quality of next-generation sequencing in clinical laboratory practice. *Nat Biotechnol.* 2012;30(11):1033-1036.
25. FDA Standards for NGS. FDA: Developing Analytical Standards for NGS Testing, workshop 12-Nov-2015. <http://www.fda.gov/downloads/MedicalDevices/NewsEvents/WorkshopsConferences/UCM468521.pdf>. Published November 12, 2015.
26. Zook JM, Chapman B, Wang J, et al. Integrating human sequence data sets provides a resource of benchmark SNP and indel genotype calls. *Nat Biotechnol.* 2014;32(3):246-251.
27. Drmanac R, Sparks AB, Callow MJ, et al. Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. *Science.* 2010;327(5961):78-81.
28. Complete Genomics. 08/2012. ftp://ftp2.completegenomics.com/vcf_files/Build37_2.0.0/. Accessed April 2015.
29. Walsh T, Casadei S, Coats KH, et al. Spectrum of mutations in BRCA1, BRCA2, CHEK2, and TP53 in families at high risk of breast cancer. *JAMA.* 2006;295(12):1379-1388.
30. Casadei S, Norquist BM, Walsh T, et al. Contribution of inherited mutations in the BRCA2-interacting protein PALB2 to familial breast cancer. *Cancer Res.* 2011;71(6):2222-2229.
31. Walsh T, Casadei S, Lee MK, et al. Mutations in 12 genes for inherited ovarian, fallopian tube, and peritoneal carcinoma identified by massively parallel sequencing. *Proc Natl Acad Sci U S A.* 2011;108(44):18032-18037.
32. Pritchard CC, Smith C, Salipante SJ, et al. ColoSeq provides comprehensive lynch and polyposis syndrome mutational analysis using massively parallel sequencing. *J Mol Diagn.* 2012;14(4):357-366.
33. Roeb W, Higgins J, King M-C. Response to DNA damage of CHEK2 missense mutations in familial breast cancer. *Hum Mol Genet.* 2012;21(12):2738-2744.
34. Pennington KP, Walsh T, Harrell MI, et al. Germline and somatic mutations in homologous recombination genes predict platinum response and survival in ovarian, fallopian tube, and peritoneal carcinomas. *Clin Cancer Res.* 2014;20(3):764-775.
35. Pennington KP, Walsh T, Lee M, et al. BRCA1, TP53, and CHEK2 germline mutations in uterine serous carcinoma. *Cancer.* 2013;119(2):332-338.
36. Antoniou AC, Casadei S, Heikkinen T, et al. Breast-cancer risk in families with mutations in PALB2. *N Engl J Med.* 2014;371(6):497-506.

Supplement

Supplemental Table 1: Known associations between genes in Color’s Hereditary Cancer Genetic Test and cancer type

Gene	Breast	Ovarian	Uterine	Colorectal	Melanoma	Pancreatic	Stomach	Prostate*
BRCA1	•	•				•		•
BRCA2	•	•			•	•		•
MLH1		•	•	•		•	•	
MSH2		•	•	•		•	•	
MSH6		•	•	•			•	
PMS2***		•	•	•				
EPCAM**		•	•	•		•	•	
APC				•		•	•	
MUTYH				•				
MITF**					•			
BAP1					•			
CDKN2A					•	•		
CDK4**					•			
TP53	•	•	•	•	•	•	•	•
PTEN	•		•	•	•			
STK11	•	•	•	•		•	•	
CDH1	•						•	
BMPRI1A				•		•	•	
SMAD4				•		•	•	
GREM1**				•				
POLD1**				•				
POLE**				•				
PALB2	•	•				•		
CHEK2	•			•				•
ATM	•					•		
NBN	•							•
BARD1	•	•						
BRIP1	•	•						
RAD51C		•						
RAD51D		•						

* Please note that research and screening guidelines for genes associated with hereditary prostate cancer are still in their early stages.
 ** Only positions known to impact cancer risk analyzed (genomic coordinates in GRCh37): CDK4: only chr12:g.58145429-58145431 (codon 24) analyzed, EPCAM: only large deletions and duplications including 3’ end of the gene analyzed, GREM1: only duplications in the upstream regulatory region analyzed, MITF: only chr3:g.70014091 (including c.952G>A) analyzed, POLD1: only chr19:g.50909713 (including c.1433G>A) analyzed, POLE: only chr12:g.133250250 (including c.1270C>G) analyzed.
 *** PMS2: Exons 12-15 not analyzed.