Somatic variant detection with Twist Oncology - DNA CGP Panel using QIAGEN® CLC Genomics Workbench

Emil Christensen[1], Anne-Mette K. Hein[1], Leif Schauser[1], Tina Han[2], James Flynn[2], Chris Pham[2], Rebecca Barnard[2]
[1] QIAGEN GmbH, Hilden, Germany
[2] Twist Bioscience, South San Francisco, USA

Introduction

The Twist Oncology - DNA CGP Panel from Twist Bioscience is designed to enable comprehensive genomic profiling (CGP) of solid tumors for pathogenic variants and scoring of microsatellite instability (MSI) and tumor mutational burden (TMB). QIAGEN CLC Genomics Workbench is a powerful analysis and visualization software with analytical tools and workflows tailored to next-generation sequencing (NGS) data, including data from a wide range of resequencing panel designs.

Here, we present highly sensitive somatic detection of commercially available tumor reference samples prepared with the Twist Oncology - DNA CGP Panel and resulting data analyzed with QIAGEN CLC Genomics Workbench.

Twist Oncology - DNA CGP Panel features

- The panel covers 2.4 Mb targeting regions of interest from 562 genes of relevance to CGP for detection of variants and scoring of MSI and TMB.
- Additional probes of 57 genes were included in order to improve the resolution of copy number variant (CNV) detection in these genes. CNV detection is coverage based, so CNVs of other genes in the panel will also be called.
- To detect known fusion events, targets were designed to investigate intronic regions of *ALK, BCR, EGFR, ETV6, EWSR1, FGFR2, FGFR3, NTRK1, NUTM1, PAX8, RET, ROS1, TMPRSS2* and *BRAF.*
- 50 targets were added for the purpose of MSI detection, bringing the total number of MSI loci used by the MSI detection tool to 183.

Methods and materials

Normal controls from Coriell Institute for Medical Research and formalin-fixed, paraffin-embedded (FFPE) samples from LGC Clinical Diagnostics (SeraCare) and Horizon Discovery® were used to evaluate workflow integrity and robustness from extracted gDNA to post-analytical bioinformatics.

Name	Vendor Cat. No.	Purpose	Reads	Average Coverage
Formalin-Compromised DNA (Moderate) Reference Standard	Horizon Discovery HD799	Demonstrate ability to call 11 SNPs, insertions, and deletions at 0.8% to 24.5% allelic frequency on moderately formalin-compromised DNA; demonstrate ability to call MSI-High	28M / 71M / 116M	597x / 754x / 1337x
Formalin-Compromised DNA (Severe) Reference Standard	Horizon Discovery HD803	Demonstrate ability to call 11 SNPs, insertions, and deletions at 0.8% to 24.5% allelic frequency on severely formalin-compromised DNA; demonstrate ability to call MSI-High	29M / 46M / 47M	531x / 438x / 529x
Mimix™ Structural Multiplex, FFPE Reference Standard	Horizon Discovery HD789	Demonstrate ability to call 9 variants with allelic frequencies ranging from 3.5% to 9.7% and CNVs at 4.5x and 8.5x MYC-N and MET focal amplifications as well as RET and ROS1 fusions; demonstrate ability to call MSI-High	41M / 166M / 117M	929x / 2194x / 1774x
Seraseq® FFPE TMB RM Score 26	SeraCare 0710-1307	Demonstrate ability to score TMB-High	142M / 132M	1733x / 1707x
Seraseq gDNA MSI-High Mix	SeraCare 0710-1670	Demonstrate ability to call MSI-High	86M / 112M	1502x / 1908x
Seraseq MSI Panel Mix AF 5% (Tumor)	SeraCare 0710-1862	Normal control for MSS and CNV baseline creation	33M	802x
Seraseq MSI Matched-WT gDNA (Normal)	SeraCare 0710-1864	Normal control for MSS and CNV baseline creation	33M	786x
HG001	Coriell NA12878	Normal control for MSS and CNV baseline creation	116M	2010x
HG002	Coriell NA24385	Normal control for MSS and CNV baseline creation	110M	1826x
HG004	Coriell NA24143	Normal control for MSS and CNV baseline creation	21M	482x

Table 1. Details of benchmarking control samples, product information, intended function and sequencing metrics in this study.

Library preparation and sequencing

Libraries were prepared with 50 ng to 100 ng of input gDNA using the Twist Library Preparation Enzymatic Fragmentation Kit 2.0 with the Twist Universal Adapter System. Pairs of 10-bp unique dual indices (UDIs) for sample barcoding were added during 6 cycles of pre-capture PCR. Pre-capture libraries of 8 samples at 300 ng each were pooled in a single tube for overnight hybridization with the Twist Oncology - DNA CGP Panel according to Twist Target Enrichment Standard Hybridization v2 workflow. Post-capture libraries were pooled and then sequenced on an MGI DNBSEQ-T7 or Illumina® NextSeq® 2000 platform, targeting at least 32M reads as PE150 (equivalent of 2000X raw coverage for 500X mean coverage).

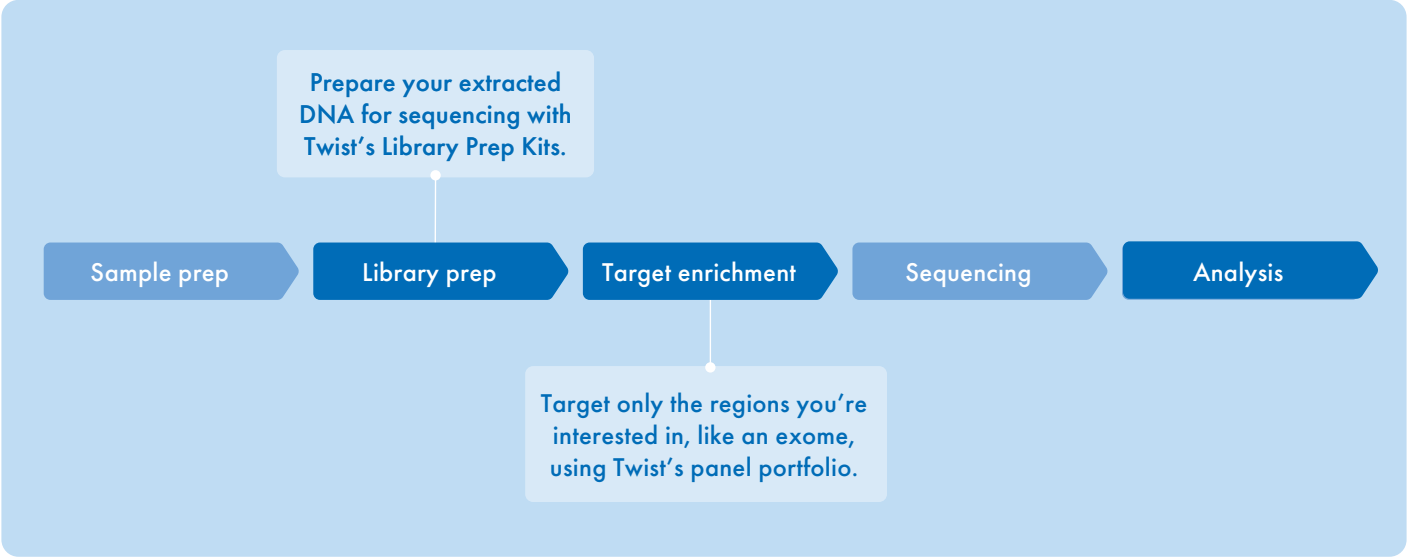


Figure 1. Schematic of the Twist Oncology - DNA CGP hybrid capture panel laboratory and analysis steps. Shown in dark blue are the topics discussed in this note (library prep, target enrichment and analysis).

Mean target coverage	515x
On-target rate	77%
Fold-80 base penalty	1.32
Duplication rate	20%
Target bases covered ≥100x	99.5%

Table 2. Relevant quality control (QC) metrics averaged across samples downsampled to 2000x raw coverage (32M 150PE reads).

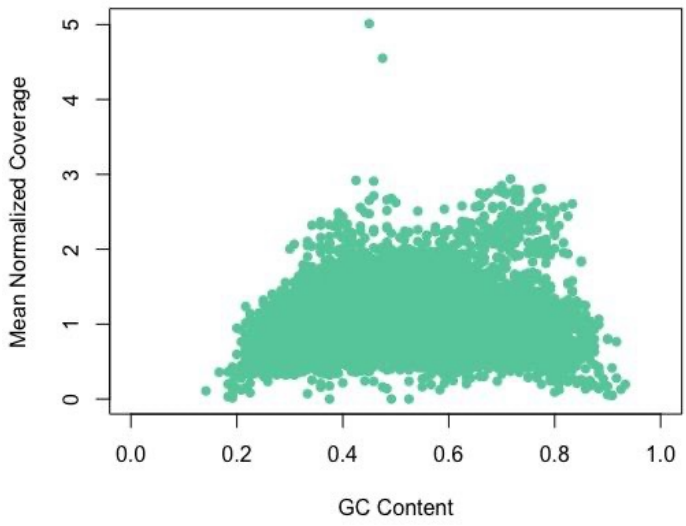


Figure 2. Normalized Twist Oncology DNA CGP panel probe coverage distribution from 2000x downsampled data across varying GC profiles.

Description of QIAGEN CLC Genomics workflow

A QIAGEN CLC Genomics Workbench "Twist Oncology - DNA CGP hybrid capture analysis" workflow was developed as a comprehensive bioinformatic pipeline for the analysis of tumor samples. It is comprised of QIAGEN CLC LightSpeed trimming, read mapping, low allele frequency variant calling with CNV and DNA fusion detection and MSI and TMB estimation steps using hg38 as the reference genome. The analysis is restricted to target regions defined by the panel, and will detect any significant variants within the targets. The workflow has thus no notion of hotspots or whitelists for detection of pre-specified variants; instead, variant detection is limited only by target regions and detection significance thresholds.

The workflow is available as a QIAGEN hosted-based service (QCI® Secondary Analysis) or as a CLC workflow for on-premise deployment.

To assess run time, fastq files of randomly selected case samples were subjected to analysis on a server-grade workstation (2x Intel® Xeon 6238R (28 core, 2.2 GHz base, 4.0 GHz boost) processors with 192 GB RAM). The total run time scaled linearly, varying from 5 minutes for a sample containing 46M reads and 20 minutes for a sample with 164M reads, which is equivalent to 7Gb and 24.6Gb as PE150 reads, respectively.

Results

Variant detection in reference samples

Detection of mutations (SNPs and Indels)

The CLC LightSpeed module provides an ultra-fast read mapping and variant calling pipeline that forms the basis of the "Twist Oncology - DNA CGP hybrid capture analysis" workflow. Variant types called directly from LightSpeed include SNPs and Indels.

Horizon reference samples HD799 and HD803 were chosen to demonstrate the workflow's ability to call 11 SNPs, insertions, and mutations present in the selected controls at a 0.8% to 24.5% allelic frequency on "moderately" and "severely" formalin-compromised DNA, respectively. Moderately compromised FFPE is defined by the manufacturer as an average fragment length of between 2,000 to 4,000 bases and DNA Integrity Number (DIN) in the range of 2.9 to 3.5. Severely compromised FFPE is defined by the manufacturer as an average fragment length <2,000 bases and a DIN in the range of 1.5 to 1.9.

For samples HD799 and HD803, all specified variants were detected at expected variant allele frequencies (VAFs) (Table 3). In addition, 15 parental cell line variants covered by the Twist Oncology - DNA CGP Panel were also detected at expected VAFs (not shown).

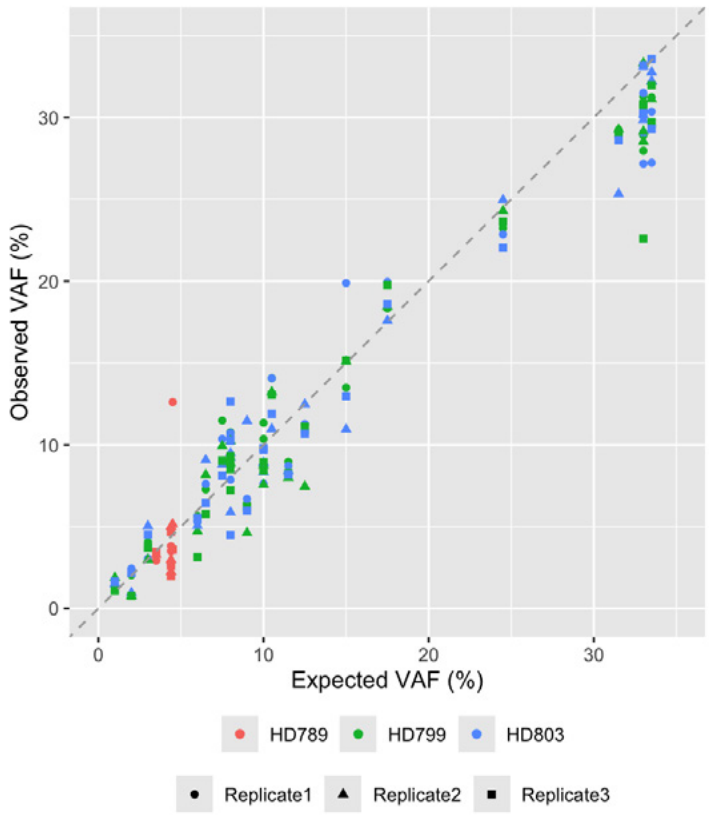
Vendor-specified variant specifications			Twist Oncology - DNA CGP detected variants in triplicate			
Gene	**Prot**	**Expected VAF (%)**	**GRCh38 coordinates**	**Mutation**	**VAF in Moderately compromised FFPE (HD799) (%)**	**VAF in Severely compromised FFPE (HD803) (%)**
NRAS	Q61K	12.50	1:114713909	G>T	10.75 / 7.45 / 11.1	11.27 / 12.47 / 10.68
PIK3CA	E545K	9.00	3:179218303	G>A	6.41 / 4.62 / 6.19	6.69 / 11.46 / 5.98
PIK3CA	H1047R	17.50	3:179234297	A>G	18.34 / 18.44 / 19.74	19.95 / 17.59 / 18.6
cKIT	D816V	10.00	4:54733155	A>T	11.34 / 7.56 / 8.42	7.62 / 8.33 / 9.7
EGFR	G719S	24.50	7:55174014	G>A	23.24 / 24.29 / 23.65	22.86 / 24.96 / 22.04
EGFR	ΔE746 - A750	2.00	7:55174773-55174787		2.03 / 0.75 / 0.77	2.43 / 0.95 / 2.16
EGFR	T790M	1.00	7:55181378	C>T	1.41 / 1.87 / 1.09	1.26 / 1.50 / 1.64
EGFR	L858R	3.00	7:55191822	T>G	4.05 / 2.97 / 3.7	3.04 / 5.04 / 4.51
BRAF	V600E	10.50	7:140753336	A>T	14.06 / 13.22 / 13.05	14.06 / 10.97 / 11.89
KRAS	G13D	15.00	12:25245347	C>T	13.49 / 15.09 / 15.15	19.88 / 10.95 / 12.97
KRAS	G12D	6.00	12:25245350	C>T	5.64 / 4.72 / 3.12	5.29 / 5.07 / 5.54

Table 3. Details of the verified variants spiked into HD799 and HD803. Vendor-specified allele frequencies and Twist Oncology - DNA CGP workflow-determined allele frequencies are indicated together with other relevant variant details.

For sample HD789, specified variants in *GNA11*, *AKT1*, and *EGFR* were detected at expected VAFs (Table 4). One exception is the mutation leading to the *PIK3CA E545K* amino acid change, which is vendor-specified at 4.5%, but detected at 12.61%, a deviation seen in only one of three replicates. The sample had fewer reads than the two other replicates (41M vs 166M and 117M), which led to insufficient coverage.

Vendor-specified variant specifications				Twist Oncology - DNA CGP detected variants in triplicate	
Gene	**Variant (AA)**	**CDS mutation**	**Expected VAF (%)**	**GRCh38 coordinates**	**Measured VAF (%)**
GNA11	Q209L	c.626A>T	4.40	19:3118944	3.82 / 5.01 / 4.67
AKT1	E17K	c.49G>A	3.50	14:104780214	2.92 / 3.27 / 3.44
PIK3CA	E545K	c.1633G>A	4.50	3:179218303	12.61 / 5.16 / 3.58
EGFR	A767_V769dup	c.2300_2308dup	4.40	7:55181317	3.49 / 2.23 / 2.59
EGFR	ΔE746 - A750	c.2235_2249del	4.40	7:55174771	2.67 / 2.97 / 1.97

Table 4. Details of the verified variants spiked into HD789. Vendor-specified VAF and Twist Oncology - DNA CGP workflow measured VAF are summarized with other relevant variant details.



All inspected variants were plotted to assess the ability of the pipeline to call VAFs in line with expectations, yielding a Pearson's correlation coefficient of 0.979 (P-value < 0.01; Figure 3).

Figure 3. Correlation of VAF in samples. Correlation between observed and expected VAF of the inspected variants in reference samples.

Detection of structural variants

Copy number variants (CNVs)

The Copy Number Variant Detection tool for targeted sequencing in QIAGEN CLC Genomics Workbench detects CNVs (gains and losses) by comparing the coverage levels in a case sample to those exhibited in a set of CNV-free controls. The tool takes a read mapping for the case sample and coverage tables or read mappings for the controls as input. A statistical assessment of the significance of the observed fold changes between the case and the control samples at the target level is made and target-level results are produced.

The Twist Oncology - DNA CGP workflow could detect CNVs at MET (4.5 copies) and *MYC-N* (8.5 copies) in the Horizon Structural Multiplex, FFPE Reference Standard (HD789). DNA from NA24143, MSI Normal, MSI 5% Tumor, NA12878 and NA24385 samples were used as controls.

The CNVs were consistently called at the expected copy numbers (Table 5). For routine testing and especially for sensitive detection of lower-fold changes, we recommend adding at least 10 control samples (processed in the same way as the test samples) to the normal control pool.

Gene	Vendor-specified copy number	Twist Oncology – DNA CGP detected copy number in triplicate
MET	4.5	5.2; 4.74; 4.86
MYC-N	8.5	9.6; 8.18; 8.34

Table 5. Structural variants. Details of the structural variants identified with the Twist Oncology - DNA CGP workflow on replicates of Horizon Structural Multiplex, FFPE Reference Standard (HD789).

Fusions

The CLC DNA Fusion detection tool takes as input a read mapping and collects evidence from broken pairs and unaligned ends to infer breakpoints and call fusion gene partners.

Twist Oncology DNA CGP has been designed to call fusions in select genes (see section Twist Oncology - DNA CGP Panel Features above). The Horizon Structural Multiplex, FFPE Reference Standard (HD789) DNA carries two fusion events: CCDC6/RET fusion at 4.6% and SLC34A2/ROS1 fusion at 9.7% AF. Both fusions were called in all three replicates of HD789 processed with the Twist Oncology - DNA CGP workflow (Table 6).

Vendor-specified fusions		Twist Oncology - DNA CGP detected fusions in three replicates		
Genes	**Expected AF (%)**	**Replicate 1 AF (%)**	**Replicate 2 AF (%)**	**Replicate 3 AF (%)**
CCDC6/RET	4.60	3.30	3.20	3.40
SLC34A2/ROS1	9.70	12.00	6.00	4.70

Table 6. Fusion allele frequency (AF) estimates. Estimated AFs identified with the Twist Oncology - DNA CGP workflow on replicates of Horizon Structural Multiplex, FFPE Reference Standard (HD789) for the two expected fusions.

MSI detection

The CLC MSI caller evaluates MSI by comparing repeat lengths at selected loci in a test sample with repeat lengths in a baseline generated with microsatellite stable (MSS) samples. At least 10 FFPE or matching tissue type is recommended for a baseline, but performance can be further enhanced with additional samples.

For MSS baseline creation, we sequenced with the Twist Oncology - DNA CGP Panel the Seraseq MSI Panel Mix AF 5% (Tumor), Seraseq® MSI Matched-WT gDNA (Normal), NA24143, NA24385, and NA12878 samples. Leave-one-out cross-validation confirmed all baseline samples as MSS. The Seraseq MSI Panel Mix AF 5% (Tumor) sample is vendor-engineered by spiking in aberrant repeat lengths for 5 "classic" MSI loci (BAT-25, BAT-26, NR-21, NR-24, and MONO-27) at 5% AF. For the CLC MSI detection algorithm, which assesses loci outlined by the tool MSIsensor2 and which do not include the "classic" loci for MSI testing, this sample represents a normal sample and is therefore useful for baseline creation.

Sample	Twist Oncology - DNA CGP Panel MSI score (% in triplicate)
HD799	High (96 / 95 / 94)
HD803	High (97 / 95 / 95)
HD789	High (93 / 93 / 93)
Seraseq gDNA MSI-High Mix	High (82 / 83)
Seraseq FFPE TMB RM Score 26	Low (4 / 3)

Table 7. MSI scores. MSI scores of samples processed with the Twist Oncology - DNA CGP workflow. Percent loci found instable when compared to stable baseline are indicated in parentheses for replicates.

Test Twist Oncology - DNA CGP-processed samples HD799, HD803, HD789, and Seracare gDNA MSI-High Mix were scored in duplicates/triplicates as MSI-High, whereas sample Seraseq FFPE TMB RM Score 26 (Seracare 0710-1307) was scored in duplicates as MSI-Low (Table 7). Percent loci called instable are highly consistent across replicates. These findings align with vendor sample descriptions or corresponding orthogonal Illumina Trusight® Oncology 500-processed scores of the same samples.

TMB scores

The CLC TMB score tallies somatic mutations per megabase of panel target region. In order to arrive at a set of somatic mutations, the germline variants have to be subtracted from variants called with the CLC LightSpeed Somatic workflow. We used dbSNP156-common and 1000-Genomes Phase 3 variants for germline filtering. Resulting TMB scores on Seraseq FFPE TMB RM Score 26 (duplicates), NA12878, NA24385, and NA24143 samples were concordant with vendor-specified/ expected scores (Table 8). The use of additional germline filters, such as GnomAD, may help reduce the scores of the Coriell samples further.

Sample	Vendor-specified / expected TMB score	Twist Oncology - DNA CGP Panel TMB score
Seraseq® FFPE TMB RM Score 26	26	26 / 25
NA12878	0-5	3
NA24385	0-5	4
NA24143	0-5	6

Table 8. TMB scores. TMB scores called on samples processed with the Twist Oncology - DNA CGP workflow.

Conclusion

The Twist Oncology - DNA CGP workflow has been designed to sensitively detect somatic variants in oncology-relevant genes and estimate MSI and TMB scores. When pairing the Twist workflow with QIAGEN CLC Genomics Workbench, we demonstrate that these biomarkers could be accurately measured in gDNA from reference samples.

Learn more about QIAGEN CLC Genomics Workbench at:
digitalinsights.qiagen.com/products-overview/discovery-insights-portfolio/qiagen-clc-genomics/

Trademarks: QIAGEN®, Sample to Insight®, QCI® (QIAGEN); Twist Bioscience®, Horizon Discovery®, Illumina®, TruSight®, NextSeq® (Illumina, Inc.); Seraseq® (SeraCare Life Sciences Inc.). Registered names, trademarks, etc. used in this document, even when not specifically marked as such, may still be protected by law.
QPRO-10934 07/2025 © 2025 QIAGEN, all rights reserved.

Ordering **qiagen.com/bioinformatics**
Technical Support **digitalinsights.qiagen.com/support**
Website **digitalinsights.qiagen.com**