




# Combining targeted sequencing and ultra-low-pass whole-genome sequencing for accurate somatic copy number alteration detection

Junfeng Fu<sup>1</sup> · Weihua Guo<sup>2</sup> · Cheng Yan<sup>2</sup> · Zhenyang Lv<sup>1</sup> · Yu Wang<sup>1</sup> · Ze Wang<sup>1</sup> · Zhe Fan<sup>1</sup> · Ting Lei<sup>1</sup> 

Received: 28 September 2020 / Revised: 14 November 2020 / Accepted: 19 January 2021 / Published online: 4 February 2021

© The Author(s), under exclusive licence to Springer-Verlag GmbH, DE part of Springer Nature 2021

## Abstract

This study investigated the feasibility of combining targeted sequencing and ultra-low-pass whole-genome sequencing (ULP-WGS) for improved somatic copy number alteration (SCNA) detection, due to its role in tumorigenesis and prognosis. Cerebrospinal fluid and matched blood samples were obtained from 29 patients with brain metastasis derived from lung cancer. Samples were subjected to targeted sequencing (genomic coverage: 300 kb) and 2×ULP-WGS. **The SCNA was detected by the CTLW\_CNV, Control-FreeC, and CNVkit methods and their accuracy was analyzed.** Eighteen tumor samples showed consistent SCNA results between the three methods, while a small fraction of samples resulted in different SCNA estimations. Further analysis indicated that consistency of SCNA highly correlated with the difference of baseline depth (normalized depth of regions without SCNA events) estimation between methods. Conflict Index showed that CTLW\_CNV significantly improved the accuracy of SCNA detection through precise baseline depth estimation. CTLW\_CNV combines targeted sequencing and ULP-WGS for improved SCNA detection. The improvement in detection accuracy is mainly due to a refined baseline depth estimation, guided by single-nucleotide polymorphism allele frequencies within the deeply sequenced region (targeted sequencing). This method is especially suitable for tumor samples with biased aneuploidy, a previously under-estimated genomic characteristic across different cancer types.

**Keywords** Somatic copy number alteration · Targeted sequencing · Ultra-low-pass whole-genome sequencing · SCNV baseline · WGD

## Introduction

Somatic copy number alterations (SCNAs) are a prevalent genotypic phenomenon closely associated with the pathogenesis of various disorders (Tang and Amon 2013; Zhang et al. 2009). As they are known to play significant roles in cancer development (Brown et al. 2017; Morikawa et al. 2018; Muñoz-Hidalgo et al. 2020), identification of SCNAs can provide important insights into the molecular basis of cancer biology (Weir et al. 2007; Zack et al. 2013).

In clinical laboratories, SCNAs are commonly identified using fluorescence in situ hybridization (FISH) or

immunohistochemistry (Soda et al. 2007). However, these experimental approaches can only query limited amount of genomic loci at a time. High-throughput techniques, such as array comparative genomic hybridization (aCGH) and single-nucleotide polymorphism (SNP) arrays, are able to search SCNAs in thousands of loci across the genome (Carter et al. 2012a, b; Maciejewski et al. 2009). With the development of advanced sequencing technologies, next-generation sequencing (NGS) has become a cost-effective alternative to investigate SCNAs at genome scale (Dong et al. 2017; Zhang and Hao 2018). Among numerous sequencing strategies, ultra-low-pass whole-genome sequencing (ULP-WGS) has been utilized for the large-scale detection of SCNAs (Adalsteinsson et al. 2017). The general approach is to segment genomic regions of both tumor and matched normal tissue, and sequencing depth of each corresponding region is normalized and compared to determine the copy number status (Xia et al. 2015a, b). However, this procedure makes biased normalization of sequencing throughput, and results in inaccurate SCNA results when dealing with tumor samples with abundant SCNA events. One such instance

✉ Ting Lei  
tuohuanyo76@163.com

<sup>1</sup> Department of Thoracic Surgery, The Second Hospital of Dalian Medical University, No.467 Zhongshan Road, Shahekou District, Dalian 116000, Liaoning Province, China

<sup>2</sup> Genetron Health (Beijing) Co. Ltd., Beijing 102206, China

is in tumors with whole-genome doubling (WGD) (Campbell et al. 2008), where genomic amplifications are widespread across >50% of the genomic region (Carter et al. 2012a, b). A 2019 pancreatic study has shown that tumors with biased aneuploidy are far more frequent than we previously expected (Priestley et al. 2019). Therefore, a reliable method is in imminent need to be able to accurately detect SCNAs in tumors with biased aneuploidy.

To achieve better SCNA detection accuracy using ULP-WGS, we presented a targeted sequencing-guided SCNA detection method (CTLW\_CNV). This was applied to cell-free DNA (cfDNA) from cerebrospinal fluid (CSF) of patients with brain metastasis derived from lung cancer. We selected these tumor samples because they are known to present prevalent SCNAs (Li et al. 2015); and aneuploidy is common in metastatic tumors (Carter et al. 2012b). The SCNA detection efficiency of CTLW\_CNV was then assessed by comparing it with two widely used algorithms: Control-FreeC (<http://bioinfo.curie.fr/projects/freec/>) (Boeva et al. 2012) and CNVkit (<https://github.com/etal/cnvkit>) (Talevich et al. 2016).

## Materials and methods

### Ethics statement

This study was approved by the ethics committee of the Second Hospital of Dalian Medical University (Dalian, China). Informed consent was obtained from each patient.

### Sample collection

A total of 29 patients with brain metastasis derived from lung adenocarcinoma were enrolled from 2017 to 2019 in the Second Hospital of Dalian Medical University. All patients underwent either MRI scans or CSF cytological examinations. In total, 10 ml of CSF obtained by lumbar puncture and 10 ml of blood was collected from each patient to perform NGS profiling.

### DNA extraction

CSF was centrifuged at 4 °C, 1900g for 5 min, and supernatants were further centrifuged at 4 °C, 16000g for 10 min. Samples were then collected and stored at −80 °C until extraction. TGuideBloodDNA Extraction Kit (TIANGEN, China) was used for gDNA extraction from blood lymphocytes according to the manufacturer's instructions. ctDNA was extracted from the CSF following the manufacturer's protocol of the MagMax<sup>TM</sup> Cell-Free DNA Isolation Kit (Thermo fisher A29319, USA). Both gDNA and ctDNA were qualified with a 2200 Bioanalyzer (Agilent Technologies, USA) in

fragment size, quality, and total concentration. The ctDNA and gDNA were used as matched tumor and normal samples respectively.

### NGS library preparation and sequencing

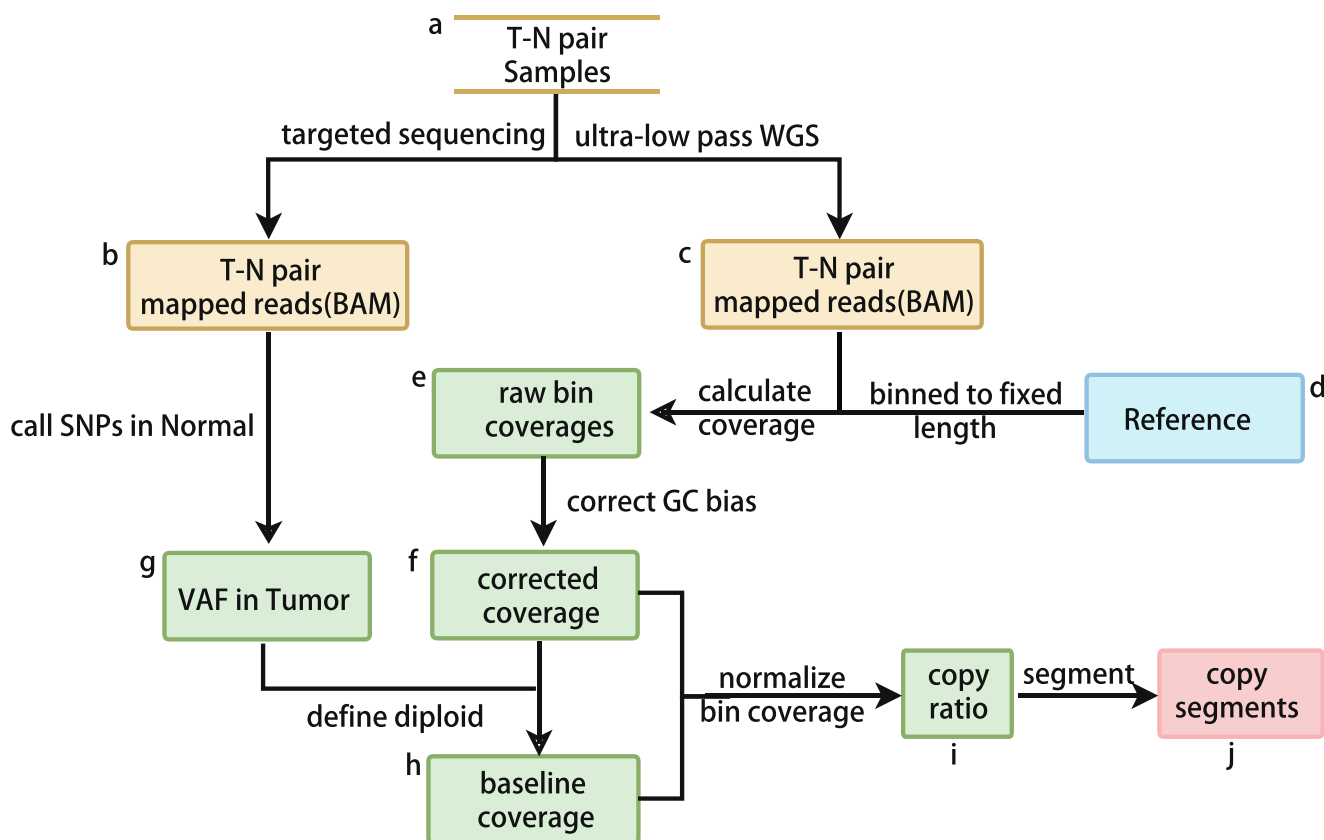
The library of gDNA was constructed by a KAPA Hyper Prep kit (Kapa Biosystems, USA) according to the manufacturer's protocols. For the library of ctDNA constructions, 4-bp random nucleotides were added as built-in tags to the both ends of ctDNA after A-tailing of the 3' ends; the adaptor was added for PCR amplification. The quantities of the library were measured with Qbit 3 (Thermo fisher, USA). These genomic barcoded DNA libraries were used as WGS libraries.

A panel of 180 genes (Genetron Health, product catalog identifier: fwa-p180-cancer) was used in our research (complete gene set was listed in Supplementary Table 1). The gDNA and ctDNA libraries obtained above were enriched for regions of this custom designed captured probe manufactured by Agilent. The 750 ng prepared libraries were hybridized with two different hybridization reagents and blocking agents in SureSelectXT Target Enrichment System (Agilent Technologies, USA). The enriched libraries were amplified with P5/P7 primer. Next, libraries were qualified by the 2200 Bioanalyzer, Qbit3, and a QPCR NGS library quantification kit (Agilent Technologies, USA).

All WGS and targeted libraries were sequenced using a 150-bp paired-end strategy on the NovaSeq system (Illumina, USA) with different coverage depths.

### Data pre-processing

The targeted sequencing and ULP-WGS data of paired tumor-normal samples were performed for data pre-processing, as follows: low-quality bases of raw data in FASTQ were trimmed by Trimmomatic-0.33 (Bolger et al. 2014) and quality control was performed using FastQC v0.11.3 (Andrews 2010). The results of quality control are shown in Supplementary Table 2. Clean reads were then aligned to the hg19 (GRCh37) reference genome using the Burrows–Wheeler Aligner (Li and Durbin 2009). Picard-tools-1.103 ("Picard Toolkit." 2019. Broad Institute, GitHub Repository. <http://broadinstitute.github.io/picard/>; Broad Institute) was used to sort the BAM file and to mark duplicated reads. The Genome Analysis Toolkit (GATK) (Depristo et al. 2011) was used for realignment and recalibration of base quality score. We got prepared mapped sequence files (bam files) after the above steps. Heterozygous single-nucleotide polymorphisms (SNPs) were detected from normal targeted bam file using GATK and variant allele frequency (VAF) at the same the locus were calculated in tumor targeted bam file using Samtools-0.1.19 (Li et al. 2009).



**Fig. 1** CTLW workflow scheme. (a) Paired tumor-normal samples were subjected to both targeted sequencing and ULP-WGS; (b, c) reads were mapped to human reference genome after quality control; (d, e, f) read counts from the mapped sequence files (bam files) of ULP-WGS were then binned into contiguous 1 Mb windows and GC bias was normalized by GATK to get coverage depth of each bin; (g) heterozygous SNPs were called from normal bam files of targeted sequencing data and variant

allele frequency at the same locus were calculated in tumor targeted bam files; (h, i) baseline coverage was defined according to VAF of heterozygous SNPs and normalized coverage of bins. The baseline coverage was then used to normalize the bin depth of WGS bam; (j) copy ratio of normalized coverage of each bin between tumor and normal sample was further segmented by CBS method to determine SCNA status of each region

## SCNA analysis by CTLW\_CNV

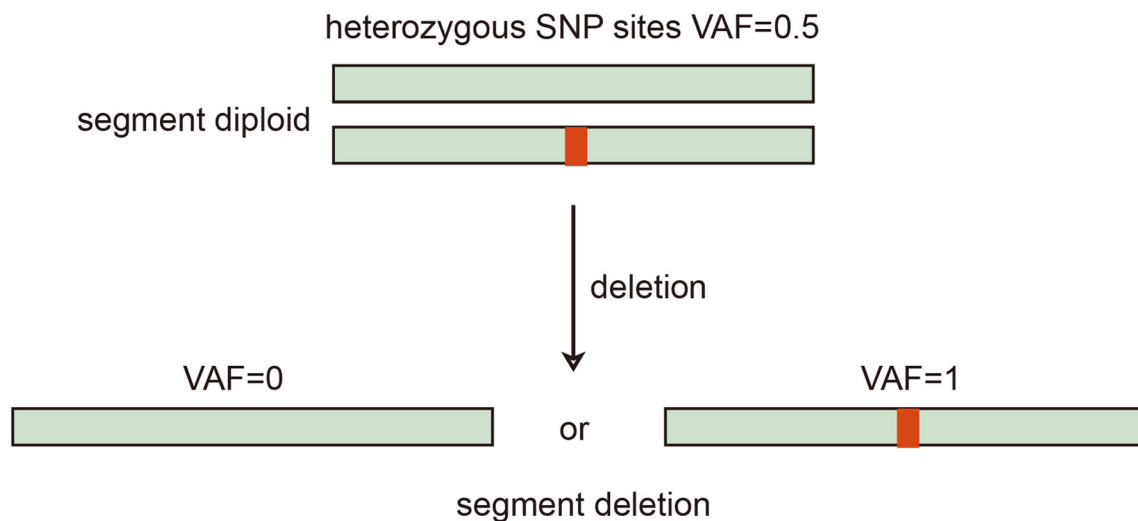
The genomic region was binned into contiguous 1 Mb windows and the corresponding sequencing depth was calculated

from the pre-processed WGS bam files. This was followed by GC bias correction using GATK. Next, we defined the baseline coverage of each tumor. We selected all genomic regions that have stable sequencing depth in both tumor and normal

**Table 1** Summary of features for SCNA algorithms

Feature	Control-FreeC	CNVkit	CTLW_CNV
Sequencing type	WGS/WES	WGS/WES/TAS	WGS
Implementation	C++/R	Python/R	Python/R
Control sample	Optional	Optional	Required
Input	SAM, BAM, pileup, bowtie, eland, arachne, psl (BLAT), BED, Eland	BAM	BAM, target SNP VAF
Graphical output	Yes	Yes	Yes
Annotation of gain/lose regions	Yes	Yes	No
GC-correction	Yes	Yes	Yes
Mappability correction	Optional	Optional	No
Contamination correction	Optional	Optional	No
Segmentation	LASSO	Optional	CBS

SCNA somatic copy number alteration, WGS whole-genome sequencing, WES whole-exome sequencing, TAS targeted amplicon sequencing, CBS circular binary segmentation



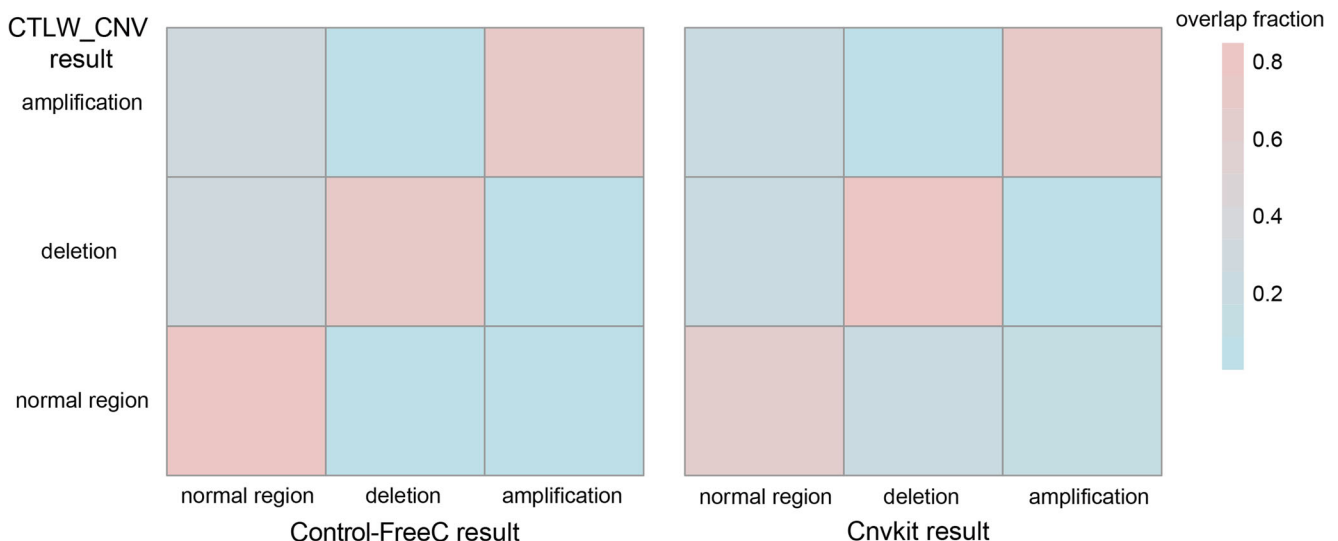
**Fig. 2** Deletion on diploid segments with heterozygous SNP sites should result in deletion segments with VAF of 0 or 1. It is Conflict SNP if the VAF of heterozygous SNP sites in the deletion region is between 40 and 60%

WGS bam as well as stable VAF of heterozygous SNP (VAF within 40~60% in both tumor and normal targeted bam). Among them, the regions of the lowest depth were chosen as diploid and the mean of bin depth in these regions was calculated as baseline coverage for each WGS bam and then used for sequencing throughput normalization. Comparing this with a traditional sequencing throughput normalization method, our method avoids over-estimation of baseline depth due to the overall high amplification of the aneuploidy genome. After normalization, the depth ratio in each genomic bin was calculated by comparing tumor to normal derived from the same patient and further log2 transformed. Lastly, SCNA status of each region was determined through circular binary segmentation (CBS) (Olshen et al. 2004) via a DNA

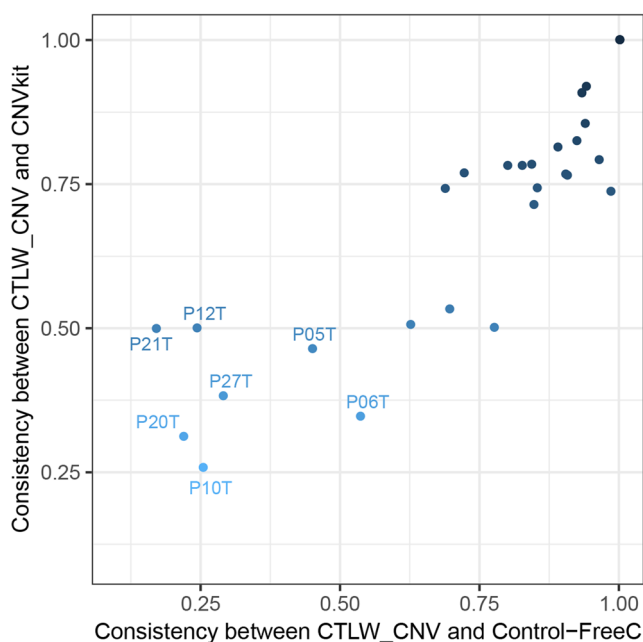
copy R package (v1.42.0). A CN segment = 2 indicated that no CNV event occurred (normal copy), >2 represented amplification, and <2 represented deletion. The parameters were set as follows: bin size = 1 M, alpha = 0.000001, and min-width = 5 for segmentation. The CTLW workflow schematic is shown in Fig. 1.

### Performance comparison between SCNA detection methods

To evaluate the accuracy of our method (CTLW\_CNV), another two algorithms (Control-FreeC and CNVkit) were used for comparison. The parameters of Control-FreeC were set as default except for window = 1000000. CNVkit parameters



**Fig. 3** The segment overlap fraction is consistent between different SCNA callers. The results of the three methods are roughly the same; for example, the Control-FreeC and CNVkit amplification segments show highly overlapping fractions on the regions CTLW\_CNV also amplified.



**Fig. 4** The consistency of somatic copy number alteration (SCNA) results between CTLW\_CNV and Control-FreeC/CNVkit across 29 samples. The darker the color, the higher the consistency. A fraction of the samples showed low SCNA consistency

were default except `-target-avg-size = 1000000`. The main features of all three methods are shown in Table 1.

### Conflict Index comparison

Given that circulating tumor DNA (ctDNA) is relatively pure in CSF, we define heterozygous SNP loci as Conflict SNP if it lays on a genomic region with a heterozygous deletion (haploid), but still has an allele frequency between 0.4 and 0.6 (Fig. 2). Correspondingly, deletion segments are defined as Conflict Segments if all the heterozygous SNPs they harbor are Conflict SNPs, and vice versa. We calculated Conflict Index (CI) as:  $CI = \# \text{ of Conflict Segments} / (\# \text{ of Conflict Segments} + \# \text{ of Non-conflict Segments})$ . The larger the CI value, the more the bias of SCNA estimation. CIs for three different algorithms across the test samples were analyzed to reflect the severity of the SCNA detection bias. The difference between CIs of CTLW\_CNV and Control-FreeC or CNVkit was analyzed with a one-sided Wilcoxon signed-rank test and a value of  $p < 0.05$  was considered statistically significant.

## Results

### CTLW\_CNV showed reliable SCNA results

We applied CTLW\_CNV, CNVkit, and Control-FreeC methods to CSF-derived cfDNA samples from 29 patients with metastatic lung cancer. To evaluate the consistency of

the SCNA detection between methods, we calculated segment fractions consistent with SCNA results from the three methods. The segment-wise comparison in Fig. 3 shows that CTLW\_CNV was generally comparable with CNVkit and Control-FreeC.

### Difference in SCNA baseline estimation contributed to inconsistent SCNA results between methods

Furthermore, we calculated sample-wise consistency between CTLW\_CNV and the other two methods. Our results show that a fraction of samples showed low SCNA consistency between CTLW\_CNV and Control-freeC and CNVkit (Fig. 4). We suspected that the inconsistency of SCNA estimation in a minority of samples was due to the difference of baseline depth estimation between methods; therefore, we plotted the baseline depths estimated by each method (Fig. 5a). As expected, the inconsistent samples in Fig. 4 also showed a greater difference in baseline depth estimation. A high correlation between SCNA estimation consistency and difference of baseline depth estimation is shown in Fig. 5b. Samples with a greater difference in CN estimation or low SCNA consistency (Fig. 5) had higher CIs. This implied a biased estimation of CN as well as SCNA.

### CTLW\_CNV exhibited more accurate SCNA estimation than Control-FreeC or CNVkit

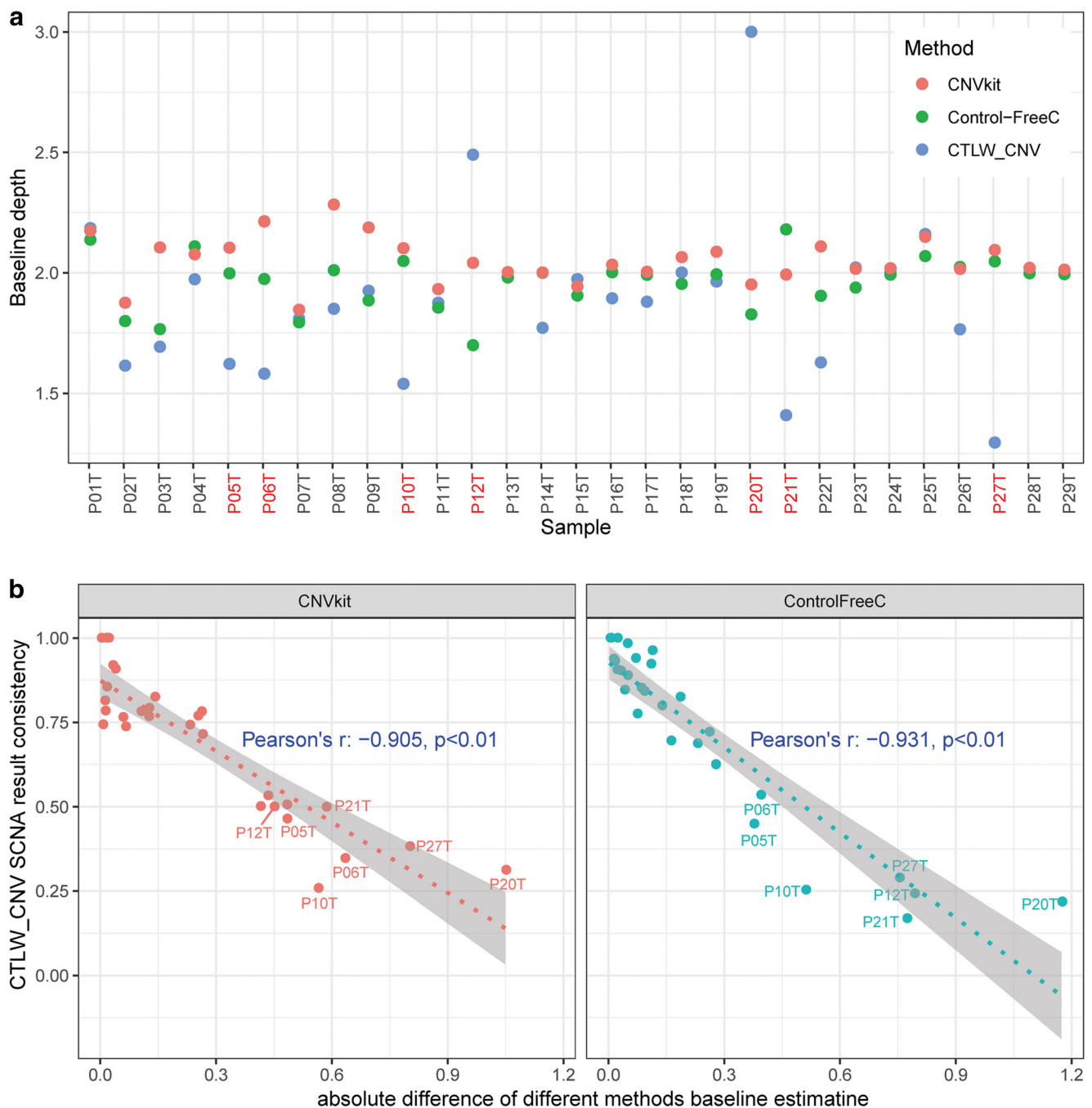
The above analysis revealed that, although the SCNA results from the three methods were roughly consistent, a small fraction of the samples showed less consistent baseline depth and SCNA estimation. We thought to compare the overall accuracy between methods based on the CI criteria. As shown in Fig. 6, CTLW\_CNV resulted in significantly lower CIs, implying improved SCNA estimation. The sample pairwise comparison in Fig. 6 revealed that the accurate SCNA estimation from CTLW\_CNV was mainly attributed by the correct estimation of baseline depth of biased aneuploidy samples.

## Discussion

Genome-wide SCNA profiling provides important information in terms of cancer development (Lee et al. 2017; Sugai et al. 2018). Numerous focal and large-scale SCNAs have been identified as tumorigenesis driving forces (Camacho et al. 2017; Luo et al. 2020). Therefore, accurate detection of SCNAs is critical to gain molecular insights into cancer biology as well as to discover potential effective therapeutic targets (Sayles et al. 2019).

CNVkit can take either WGS or targeted sequencing data as input for SCNA detection. Different mode uses either on-target reads or off-target reads to infer copy number of



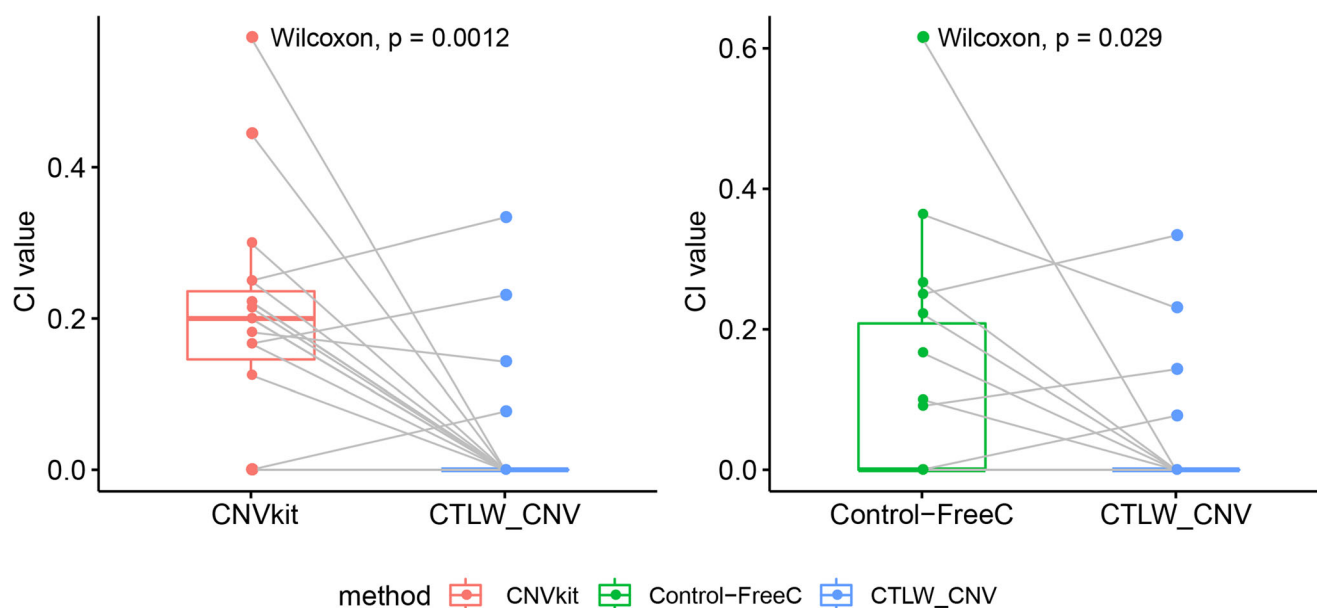


**Fig. 5** **a** The average read depth of the stable segment (baseline) of each sample analyzed by three methods. It varies widely in some samples. **b** The difference between the baseline estimation of CTLW\_CNV and the

other methods led to the different SCNA results which showed high Pearson correlation coefficients and significant  $p$  values

corresponding genomic regions and median-centers the corrected copy ratios when normalize test samples to the reference (Talevich et al. 2016). Here, we developed CTLW\_CNV by combining targeted sequencing and ULP-WGS, aiming to achieve more accurate genome-wide SCNA detection via different baseline estimation approach. The normalization of sequencing depth has been the major challenge for accurate SCNA detection. By combining targeted sequencing and ULP-WGS, better normalization was achieved

by defining stable genomic regions using coverage depth information derived from WGS data and heterozygous SNP allele frequency information derived from targeted sequencing data. Our approach avoided biased normalization of sequencing throughput specially when dealing with tumor samples with abundant SCNA events. We also introduced CI to measure the performance of different SCNA calling methods. CI reflects the conflict of heterozygous SNPs within region of haplotype (loss of one copy) inferred by given method. The



**Fig. 6** Comparison of Conflict Index (CI) between CTLW\_CNv and the other two methods by one-sided Wilcoxon signed-rank test. CIs of CTLW\_CNv are significantly lower, which means less SCNA error

larger the CI value, the more bias of SCNA estimation. We calculated CIs of SCNA estimated on testing samples by different methods. One-sided Wilcoxon signed-rank test showed that CTLW\_CNv owned significantly lower CI, indicating higher overall accuracy in SCNA detection largely due to its unbiased estimation of baseline depth.

In addition, tumor cell ploidy, the average copy number of a cancer genome, has a significant impact on SCNA detection (Zack et al. 2013). Aneuploid genomes with chromosomal gains and losses have been observed in more than 70% of cancers (Nagaoka et al. 2012). SCNAs caused by aneuploidy affect the expression of cancer-related genes and downstream oncogenic pathways (Ben-David et al. 2014). Aneuploidy can also induce genome instability and consequently accelerate tumor cell development (Pavelka et al. 2010). Moreover, aneuploidy and SCNA levels in cancers have been shown to positively correlate with cell proliferation, while they negatively associate with immune cell infiltration and patient survival in immunotherapy trials (Buccitelli et al. 2017; Davoli et al. 2017; Taylor et al. 2018). For tumor samples with biased aneuploidy, SCNA identification based on sequencing depth alone tends to over- or under-estimate the copy number, by incorrectly assessing baseline depth (Zhang and Hao 2015). In this study, we evaluated the baseline depth of three methods. Expectedly, the consistency of SCNA status between methods was negatively correlated with the difference of baseline estimation. We therefore introduced the allele frequency of SNPs from targeted sequencing for improved baseline depth estimation and consequently determined more precise SCNA detection.

Our analysis demonstrated that CTLW\_CNv was able to improve the accuracy of SCNA detection by combining targeted

sequencing and ULP-WGS. The improvement is more significant when dealing with tumor samples with biased aneuploidy. However, one drawback of CTLW\_CNv is that a high fraction of “normal tissue contamination” in a low purity sample may affect the baseline depth estimation and consequently compromise SCNA detection. Further analysis is required to define the correlation between tumor purity and SNP allele frequency changes in diploid cells. Such correlation can further help construct purity-adjusted baseline depth estimations. Lastly, it is worth mentioning that it is not our intention to propose using both targeted sequencing and ULP-WGS in routine NGS experiments, but rather as a complementary approach to evaluate the heterogeneity of aneuploidy in tumors.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s10142-021-00767-y>.

**Author contribution** Conception and design of the research: TL, JF, WG, and CY; acquisition of data: HF, ZL, YW, ZW, and ZF; analysis and interpretation of data: JF, WG, and CY; statistical analysis: WG and CY; drafting the manuscript: JF; revision of manuscript for important intellectual content: TL, WG, and CY. All authors read and approved the final manuscript.

**Data availability** All data generated or analyzed during this study are included in this published article.

## Declarations

**Ethics approval and consent to participate** This study was approved by the ethics committee of The Second Hospital of Dalian Medical University (Dalian, China). Informed consent was obtained from each patient.

**Consent for publication** Not applicable.

**Conflict of interest** The authors declare no competing interests.

## References

- Adalsteinsson VA, Ha G, Freeman SS, Choudhury AD, Stover DG, Parsons HA, Gydush G, Reed SC, Rotem D, Rhoades J (2017) Scalable whole-exome sequencing of cell-free DNA reveals high concordance with metastatic tumors. *Nat Commun* 8:1–13
- Andrews S (2010) FastQC: a quality control tool for high throughput sequence data. Babraham Bioinformatics. Babraham Institute, Cambridge
- Ben-David U, Arad G, Weissbein U, Mandefro B, Maimon A, Golan-Lev T, Narwani K, Clark AT, Andrews PW, Benvenisty N (2014) Aneuploidy induces profound changes in gene expression, proliferation and tumorigenicity of human pluripotent stem cells. *Nat Commun* 5:1–11
- Boeva V, Popova T, Bleakley K, Chiche P, Cappel J, Schleiermacher G, Janoueix-Lerosey I, Delattre O, Barillot E (2012) Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data. *Bioinformatics* 28:423–425
- Bolger A, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120
- Brown D, Smeets D, Székely B, Larmont D, Szász AM, Adnet P-Y, Rothé F, Rouas G, Nagy ZI, Faragó Z (2017) Phylogenetic analysis of metastatic progression in breast cancer using somatic mutations and copy number aberrations. *Nat Commun* 8:1–13
- Buccitelli C, Salgueiro L, Rowald K, Sotillo R, Mardin BR, Korbel JO (2017) Pan-cancer analysis distinguishes transcriptional changes of aneuploidy from proliferation. *Genome Res* 27:501–511
- Camacho N, Van Loo P, Edwards S, Kay JD, Matthews L, Haase K, Clark J, Dennis N, Thomas S, Kremeyer B (2017) Appraising the relevance of DNA copy number loss and gain in prostate cancer using whole genome DNA sequence data. *PLoS Genet* 13: e1007001
- Campbell PJ, Stephens PJ, Pleasance ED, O'Meara S, Li H, Santarius T, Stebbings LA, Leroy C, Edkins S, Hardy C, Teague JW, Menzies A, Goodhead I, Turner DJ, Clee CM, Quail MA, Cox A, Brown C, Durbin R, Hurler ME, Edwards PA, Bignell GR, Stratton MR, Futreal PA (2008) Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing. *Nat Genet* 40:722–729
- Carter SL, Cibulskis K, Helman E, McKenna A, Shen H, Zack T, Laird PW, Onofrio RC, Winckler W, Weir BA (2012a) Absolute quantification of somatic DNA alterations in human cancer. *Nat Biotechnol* 30:413–421
- Carter SL, Cibulskis K, Helman E, McKenna A, Shen H, Zack T, Laird PW, Onofrio RC, Winckler W, Weir BA, Beroukhim R, Pellman D, Levine DA, Lander ES, Meyerson M, Getz G (2012b) Absolute quantification of somatic DNA alterations in human cancer. *Nat Biotechnol* 30:413–421
- Davoli T, Uno H, Wooten EC, Elledge SJ (2017) Tumor aneuploidy correlates with markers of immune evasion and with reduced response to immunotherapy. *Science* 355:eaaf8399
- Depristo MA, Banks E, Poplin R, Garimella KV, Maguire J, Hartl C, Philippakis AA, Angel GD, Rivas MA, Hanna M (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 43:491–498
- Dong Z, Xie W, Chen H, Xu J, Wang H, Li Y, Wang J, Chen F, Choy KW, Jiang H (2017) Copy-number variants detection by low-pass whole-genome sequencing. *Curr Protoc Human Genet* 94:8.17. 11–8.17. 16
- Lee S, Lee J, Sim SH, Lee Y, Moon KC, Lee C, Park W-Y, Kim NK, Lee S-H, Lee H (2017) Comprehensive somatic genome alterations of urachal carcinoma. *J Med Genet* 54:572–578
- Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25:1754–1760
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth GT, Abecasis GR, Durbin R (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* 25:2078–2079
- Li F, Sun L, Zhang S (2015) Acquisition of DNA copy number variations in non-small cell lung cancer metastasis to the brain. *Oncol Rep* 34:1701–1707
- Luo H, Xu X, Yang J, Wang K, Wang C, Yang P, Cai H (2020) Genome-wide somatic copy number alteration analysis and database construction for cervical cancer. *Molec Genet Genom* 1–9
- Maciejewski JP, Tiu RV, O'Keefe C (2009) Application of array-based whole genome scanning technologies as a cytogenetic tool in hematological malignancies. *Br J Haematol* 146:479–488
- Morikawa A, Hayashi T, Kobayashi M, Kato Y, Shirahige K, Itoh T, Urashima M, Okamoto A, Akiyama T (2018) Somatic copy number alterations have prognostic impact in patients with ovarian clear cell carcinoma. *Oncol Rep* 40:309–318
- Muñoz-Hidalgo L, San-Miguel T, Megías J, Monleón D, Navarro L, Roldán P, Cerdá-Nicolás M, López-Ginés C (2020) Somatic copy number alterations are associated with EGFR amplification and shortened survival in patients with primary glioblastoma. *Neoplasia* 22:10–21
- Nagaoka SI, Hassold TJ, Hunt PA (2012) Human aneuploidy: mechanisms and new insights into an age-old problem. *Nat Rev Genet* 13: 493–504
- Olshen AB, Venkatraman E, Lucito R, Wigler M (2004) Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* 5:557–572
- Pavelka N, Rancati G, Li R (2010) Dr Jekyll and Mr Hyde: role of aneuploidy in cellular adaptation and cancer. *Curr Opin Cell Biol* 22:809–815
- Priestley P, Baber J, Lolkema MP, Steeghs N, de Bruijn E, Shale C, Duyvesteyn K, Haidari S, van Hoeck A, Onstenk W (2019) Pan-cancer whole-genome analyses of metastatic solid tumours. *Nature* 575:210–216
- Sayles LC, Breese MR, Koehne AL, Leung SG, Lee AG, Liu H-Y, Spillinger A, Shah AT, Tanasa B, Straessler K (2019) Genome-informed targeted therapy for osteosarcoma. *Cancer Discov* 9:46–63
- Soda M, Choi YL, Enomoto M, Takada S, Yamashita Y, Ishikawa S, Fujiwara S, Watanabe H, Kurashina K, Hatanaka H (2007) Identification of the transforming EML4–ALK fusion gene in non-small-cell lung cancer. *Nature* 448:561–566
- Sugai T, Takahashi Y, Eizuka M, Sugimoto R, Fujita Y, Habano W, Otsuka K, Sasaki A, Yamamoto E, Matsumoto T (2018) Molecular profiling and genome-wide analysis based on somatic copy number alterations in advanced colorectal cancers. *Mol Carcinog* 57:451–461
- Talevich E, Shain AH, Botton T, Bastian BC (2016) CNVkit: genome-wide copy number detection and visualization from targeted DNA sequencing. *PLoS Comput Biol* 12:e1004873
- Tang Y-C, Amon A (2013) Gene copy-number alterations: a cost-benefit analysis. *Cell* 152:394–405
- Taylor AM, Shih J, Ha G, Gao GF, Zhang X, Berger AC, Schumacher SE, Wang C, Hu H, Liu J (2018) Genomic and functional approaches to understanding cancer aneuploidy. *Cancer Cell* 33:676–689. e673
- Weir BA, Woo MS, Getz G, Perner S, Ding L, Beroukhim R, Lin WM, Province MA, Kraja AT, Johnson LA (2007) Characterizing the cancer genome in lung adenocarcinoma. *Nature* 450:893–898
- Xia S, Huang C-C, Le M, Dittmar R, Du M, Yuan T, Guo Y, Wang Y, Wang X, Tsai S (2015a) Genomic variations in plasma cell free DNA differentiate early stage lung cancers from normal controls. *Lung Cancer* 90:78–84
- Xia S, Kohli M, Du M, Dittmar RL, Lee A, Nandy D, Yuan T, Guo Y, Wang Y, Tschannen MR (2015b) Plasma genetic and genomic



- abnormalities predict treatment response and clinical outcome in advanced prostate cancer. *Oncotarget* 6:16411–16421
- Zack TI, Schumacher SE, Carter SL, Cherniack AD, Saksena G, Tabak B, Lawrence MS, Zhang C, Wala J, Mermel CH (2013) Pan-cancer patterns of somatic copy number alteration. *Nat Genet* 45:1134–1140
- Zhang Z, Hao K (2015) SAAS-CNV: a joint segmentation approach on aggregated and allele specific signals for the identification of somatic copy number alterations with next-generation sequencing data. *PLoS Comput Biol* 11:e1004618
- Zhang Z, Hao K (2018) Using SAAS-CNV to detect and characterize somatic copy number alterations in cancer genomes from next generation sequencing and SNP array Data. In: Bickhart DM (ed) *Copy Number Variants*. Springer, New York, pp 29–47
- Zhang F, Gu W, Hurles ME, Lupski JR (2009) Copy number variation in human health, disease, and evolution. *Annu Rev Genomics Hum Genet* 10:451–481
- Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.