Supplementary material for "Multiresolution correction of GC bias and application to identification of copy number alterations"

Ho Jang[1] **and Hyunju Lee**[1,*]

[1]School of Electrical Engineering and Computer Science, Gwangju Institute of Science and Technology, Gwangju 500-712, South Korea
[*]hyunjulee@gist.ac.kr

ABSTRACT

This is a supplementary material for the "Multiresolution correction of GC bias and application to identification of copy number alterations."

Contents

1 Methods

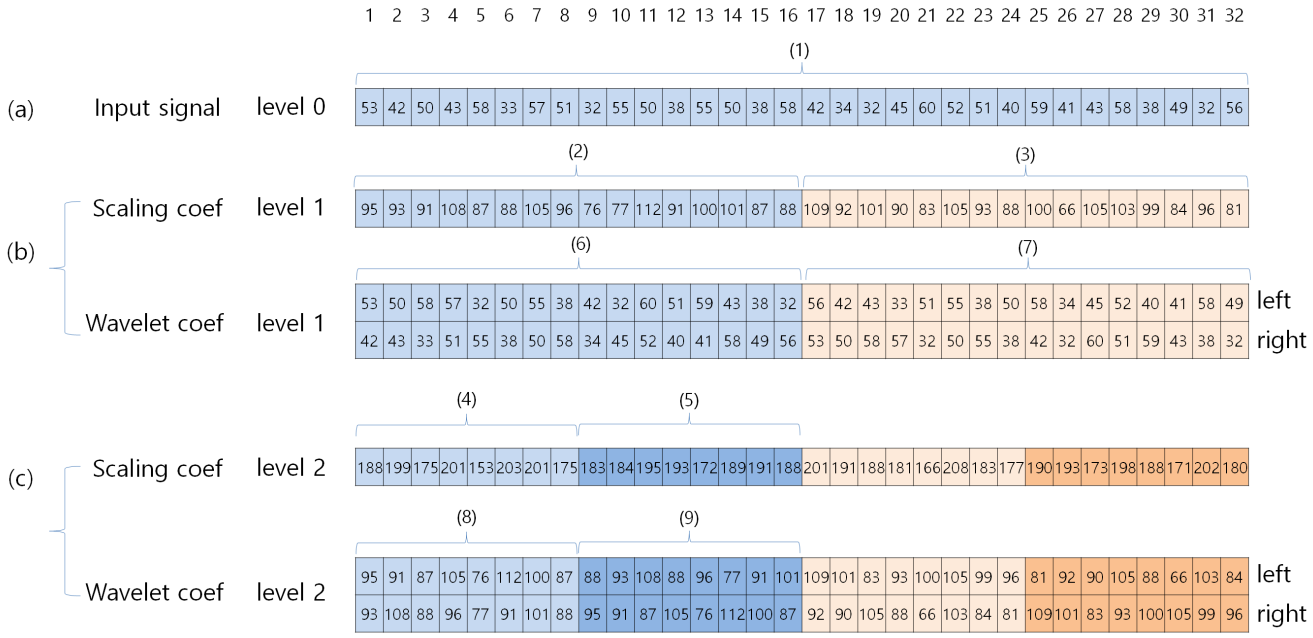

Figure S1. Wavelet and scaling coefficients and their original locations.

1.1 Modeling the relation between a scaling coefficient and its GC proportion

Figure S1 shows an example of GC proportion calculation. The GC proportion for level 1 scaling coefficients is calculated from the sum of the GC proportions of the two neighboring input bins [Figure S1 (1) to Figure S1 (2)]. In the case of the GC proportion from a shifted input signal, the GC proportion of the scaling coefficient is the sum of the two GC proportions of the shifted input signal [the shifted signal of Figure S1 (1) to Figure S1 (3)]. The GC proportion of level 2 scaling coefficients is the sum of the level 1 GC proportions previously calculated [Figure S1 (2) to Figure S1 (4)]. In the case of the level 2 GC proportion of the scaling coefficients from the shifted level 1 scaling coefficients, the GC proportion of the scaling coefficient is the sum of the two GC proportions of shifted level 1 scaling coefficients [the shifted signal of Figure S1 (2) to Figure S1 (5)].

1.2 Modeling the relation between a wavelet coefficient and its GC proportion

Figure S1 presents an example of GC proportion calculation. The GC proportions for level 1 wavelet coefficients are extracted from the GC proportions of the two neighboring input bins [Figure S1 (1) to Figure S1 (6)]. In the case of the GC proportions from the shifted input signal, the GC proportions of the wavelet coefficient are extracted from the two GC proportions of the shifted input signal [the shifted signal of Figure S1 (1) to Figure S1 (7)]. The GC proportions of level 2 wavelet coefficients come from the GC proportions of the level 1 scaling coefficients [Figure S1 (2) to Figure S1 (8)]. In the case of the level 2 GC proportions of the wavelet coefficients from the shifted level 1 scaling coefficients, the GC proportions come from the two GC proportions of shifted level 1 scaling coefficients [the shifted signal of Figure S1 (2) to Figure S1 (9)].

1.3 An example of a relation between coefficients and their GC proportion

In Figure 3 of the main text, we gave an example of a fitted scaling coefficient and wavelet coefficient from simulated GC-biased sequencing reads. The scaling coefficient in the figure is left skewed. We give another example to help understand the relation between the coefficients and GC proportion. Figure S2 depicts the relation between fitted coefficients and the GC proportion from symmetrically biased simulated sequencing data. We used Pysim-sv to generate simulated reads and simulate GC bias, on the basis of the pattern from the our custom formula $y = -20 \times (x - 0.5)^2 + 1$, where x represents the GC proportion of the sequencing reads, and y denotes a sampling rate. Figure S2 (a) shows the distribution of raw scaling coefficient values and their GC proportions at decomposition level 1. The red curve is the LOESS-fitted scaling coefficients. The curve has a symmetrical shape centering around the GC proportion of 0.5. Figure S2 (b) illustrates the smoothed results representing the relation between GC proportions of consecutive genomic regions and wavelet coefficients at decomposition level 1. The x-axis represents the left-hand genomic region, y-axis denotes the right-hand genomic region, and the color represents the values of wavelet coefficients. In this figure, when the GC proportion of the left-hand region is $\sim$0.5 and the GC proportion of the right-hand region is far from 0.5, the value of the wavelet coefficient between the two neighboring regions is negative (blue pixel) and DOC decreases.

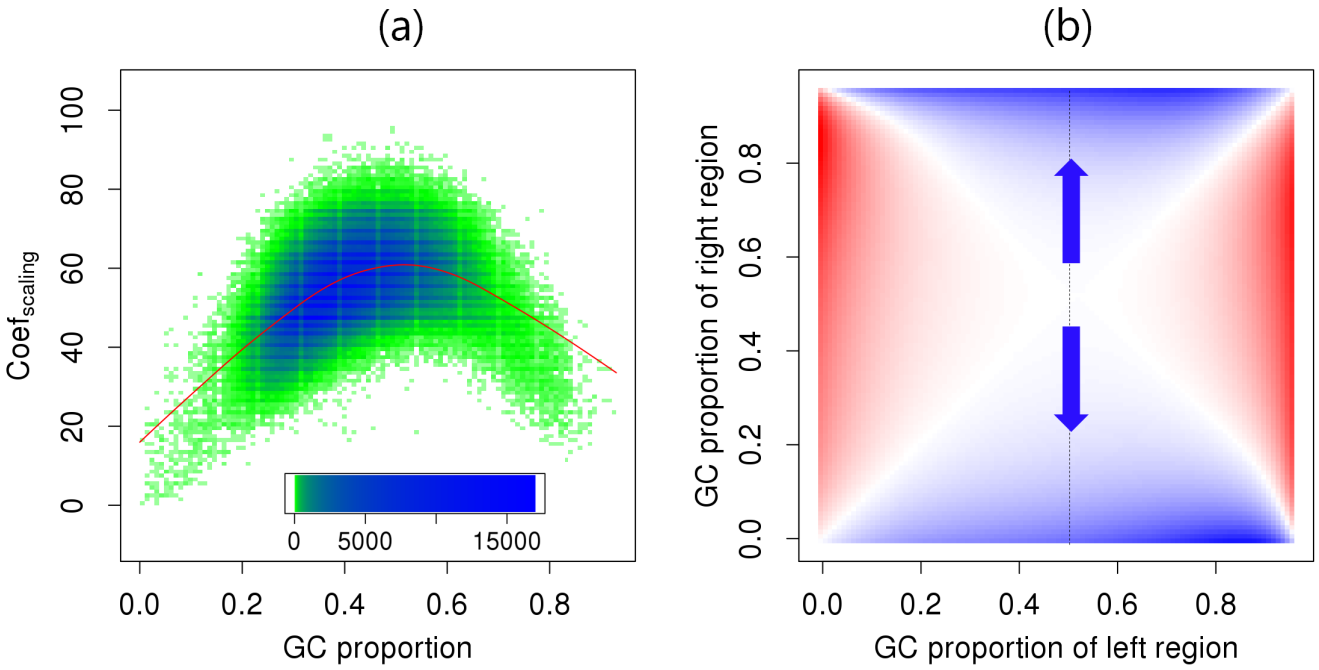

Figure S2. The relation between coefficients and their GC proportions of the simulated GC-biased sequencing reads. (a) The relation between the GC proportion and scaling coefficient. The heatmap illustrates the distribution of raw scaling coefficient values. The x-axis means the GC proportion of the coefficients, and the y-axis denotes the values of the scaling coefficients. Most of scaling coefficients are concentrated in blue areas. The scaling coefficients are rarely distributed to the green areas. The red curve is the LOESS-fitted scaling coefficient depending on the specific GC proportion. (b) The relation between two neighboring GC proportions and the wavelet coefficient. The heatmap depicts the relation between the values of wavelet coefficients (pixel color) and the GC proportion of two neighboring genomic regions (x-axis and y-axis) from 2D kernel smoothing. Red areas represent increasing DOC values of the right-hand genomic region compared to the left-hand region in the two adjacent genomic regions. The blue areas indicate the opposite case.

1.4 Stopping criteria for determining the decomposition level

In Figure 9 of the main text, we demonstrated that we can make GC fluctuation patterns in DOC_{GC} similar to the pattern in DOC_{raw} as we increase the decomposition level. If we correct GC bias by dividing DOC_{raw} by DOC_{GC}, we can see a reduction of GC bias in the CN_{Wabico} signal in Figure 8 of the main text. GC curves in Figure 8 also represent severity of GC bias in the signal. If the shape of the GC curve of the DOC signal is closer to the horizontal line, it may indicate that the effect of GC bias is smaller. In short, the increase in the decomposition level may reduce GC bias.

Nonetheless, throughout the analyses, we found that in some cases, a high decomposition level does not always yield good results on GC bias correction. In other cases, even if performance is slightly improved, the improvement is insignificant in relation to the computation required. Thus, the proper decomposition level has to be determined manually by the user or by the specific criteria.

Consequently, in Wabico, we use the GC curve to determine the proper decomposition level. Here, we tried to measure how close the shape of the curve is to the horizontal line. For this task, we designed a criterion based on the amount of changes between GC curves of the current level decomposition and the decomposition of the next level. In Supplementary Figure S3 (a), GC curves of CN_{Wabico} from the decomposition levels 1 and 2 are shown. The curve was changed from level 1 to level 2. Some part of the curve moved up and the other part of the curve moved down, and thus the shape of the curve became closer to the horizontal line. In Supplementary Figure S3 (b), the GC curve from decomposition level 3 is presented, and the shapes of the curves of levels 2 and 3 are compared. The shape of the curve became closer to the horizontal line than before. On the other hand, for some parts of the curve, their direction of change is inconsistent with the initial direction of change in Figure S3 (a) (marked by X below the curve). For example, in Supplementary Figure S3 (a), the third arrow moved down, whereas in Figure S3 (b), the third arrow moved up. Thus, we set the ratio of inconsistency in the change of GC curves as the threshold for stopping the decomposition. We also set a specific maximum decomposition level for stopping decomposition. In our analyses, if the decomposition level exceeded 10, the GC correction effect seemed to be insignificant in many samples. In conclusion, we stop signal decomposition according to the inconsistency ratio and a specific maximum decomposition level.

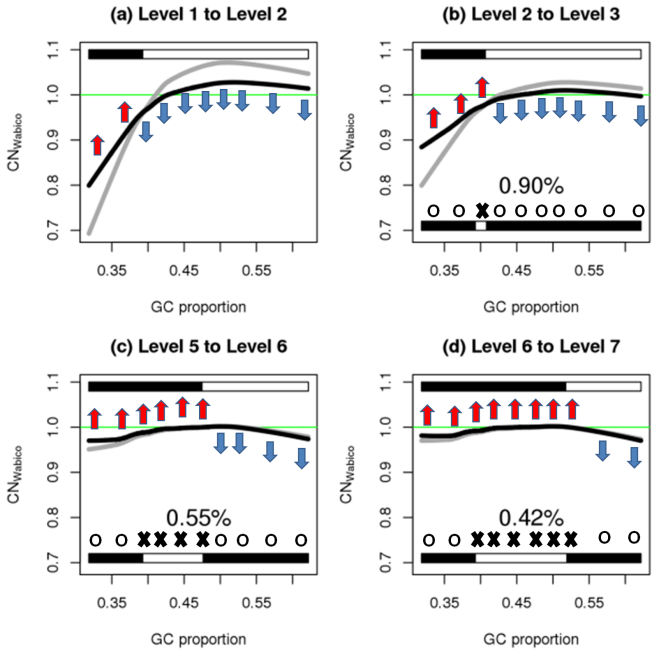

Figure S3. An example of multiresolution GC bias correction. (a) CN_{Wabico} values from level 1 (gray) to level 2 (black). (b) CN_{Wabico} values from level 2 (gray) to level 3 (black). (c) CN_{Wabico} values from level 5 (gray) to level 6 (black). (d) CN_{Wabico} values from level 6 (gray) to level 7 (black).

In Figure S3, we explained effects of decomposition level changes using GC curves. Figure S3 further explains effects of decomposition levels, and inconsistency between the initial direction of change and the direction of change of CN values across genomic markers. Figure S4 (a) shows raw CN signals only with GC biases, but without CN alteration events across the genome. We assume that the CN signals of left and right genomic parts in the figure should decrease while the CN signals of the middle part in the figure should increase. Figure S4 (b) shows the decrease of the fluctuation in CN signal as Wabico is applied up to the decomposition level N. Figure S4 (c) shows the results of corrected CN signals as it is applied up to the level N+1. Although CN signal values of some genomic markers consistently change with the previous direction of change (blue arrows), there exist genomic markers whose direction of change of the CN value is inconsistent with the previous direction of change of the CN values (red arrows). As shown in Figure S4 (c), because the increase of inconsistencies does not always guarantee the proper GC bias correction, the level of inconsistency was used as an option for choosing proper decomposition level.

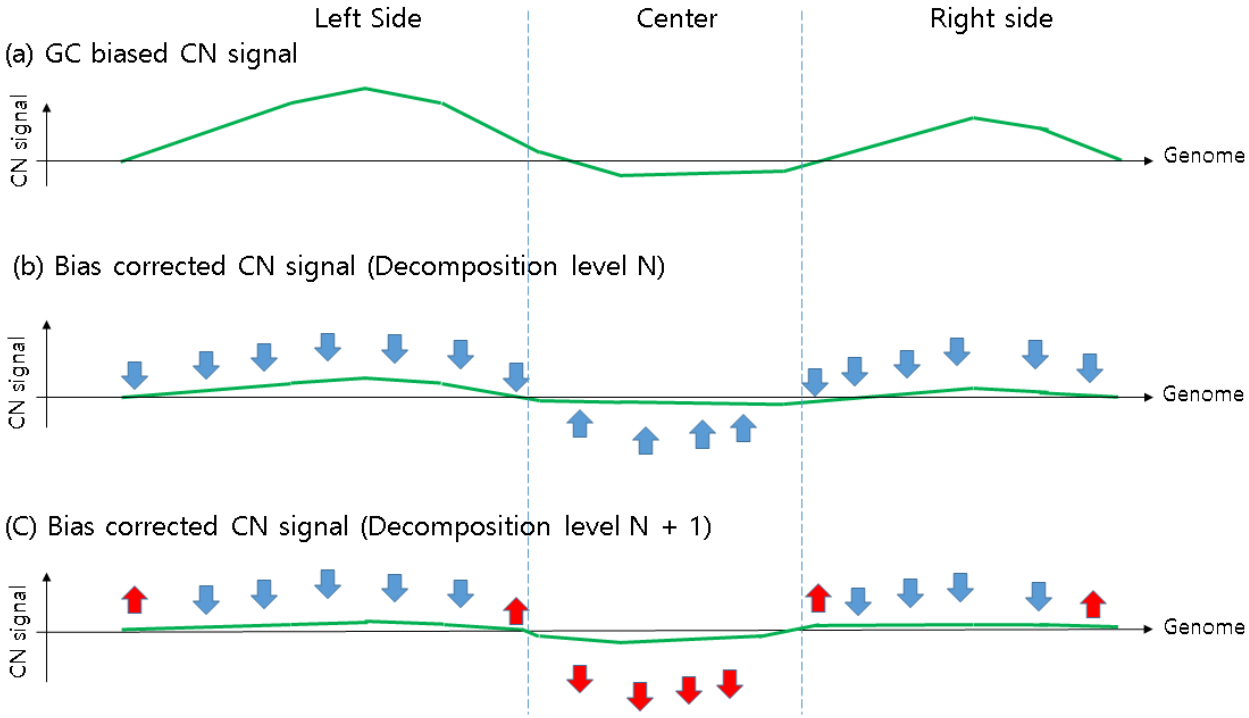

Figure S4. *CN* value changes across genomic markers.

2 Results

2.1 The effect of multiresolution decomposition in simulation data

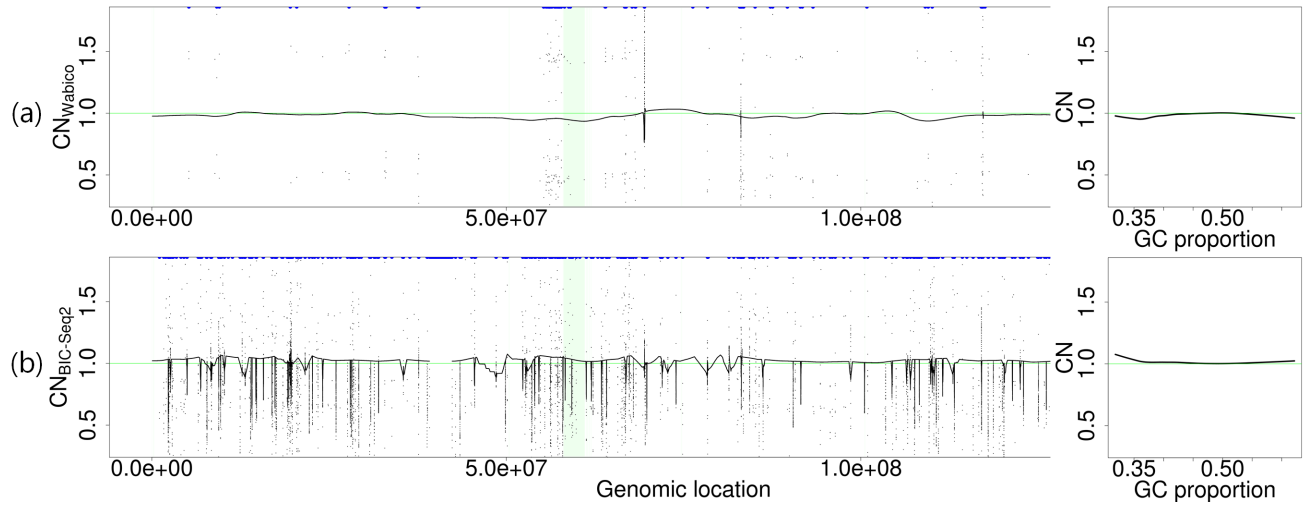

Figure S5. Copy number (*CN*) ratio signals of simulated reads as corrected by Wabico and BIC-Seq2–based expected read counts.

2.2 Identification of cancer genes in WGS datasets

Given denoised CN_{Wabico} signals, we assume that a cancer gene is altered if the following two conditions are satisfied.

1. The length of *CN* alteration including the cancer gene is less than 25% of the chromosome arm.
2. The difference between the *CN* ratio of the gene and its neighboring *CN* ratios was greater than 0.1. When the difference was less than 0.1 owing to other biases such as contamination with normal cells, we decided whether the cancer gene is altered based on y^*_{HIGH} values used in our previous study.[1] Positive y^*_{HIGH} values around the cancer gene indicate that the cancer gene is focally amplified, and negative y^*_{HIGH} values mean that the gene is focally deleted. To generate y^*_{HIGH}, we employed the same parameters as in our previous study.[1]

2.3 Identification of *CN* variations in simulated WGS data

In Figure 7 of the main text, we mainly demonstrated that Wabico shows stabler GC bias correction performance as compared to other methods on a set of simulated sequences at various levels of GC bias. In this section, we generated another set of simulated sequences with various levels of GC bias and 10 *CN* variations that were evenly spaced (Figure S6 (a)). We compared the *CN* ratios of the signals denoised by CN_{Wabico} and $CN_{BIC-Seq2}$. We determined whether the identified genomic regions from simulated data are true positive (TP), false positive (FP), true negative (TN), and false negative (FN) by means of thresholds. If the value of *CN* yielded by each method was greater or less than a threshold for amplification or deletion for the region of alteration, respectively, then the region was considered TP. Figure S6 (b) shows F1 scores for various thresholds. The F1 scores of CN_{Wabico} (black color) were consistently higher than those of $CN_{BIC-Seq2}$ (blue color) at any GC bias severity (z values). Note that one *CN* deletion event from 49,235,360 base pair to 49,335,359 base pair was not identified by any methods because there are few markers having uniquely mappable positions in this genomic region.

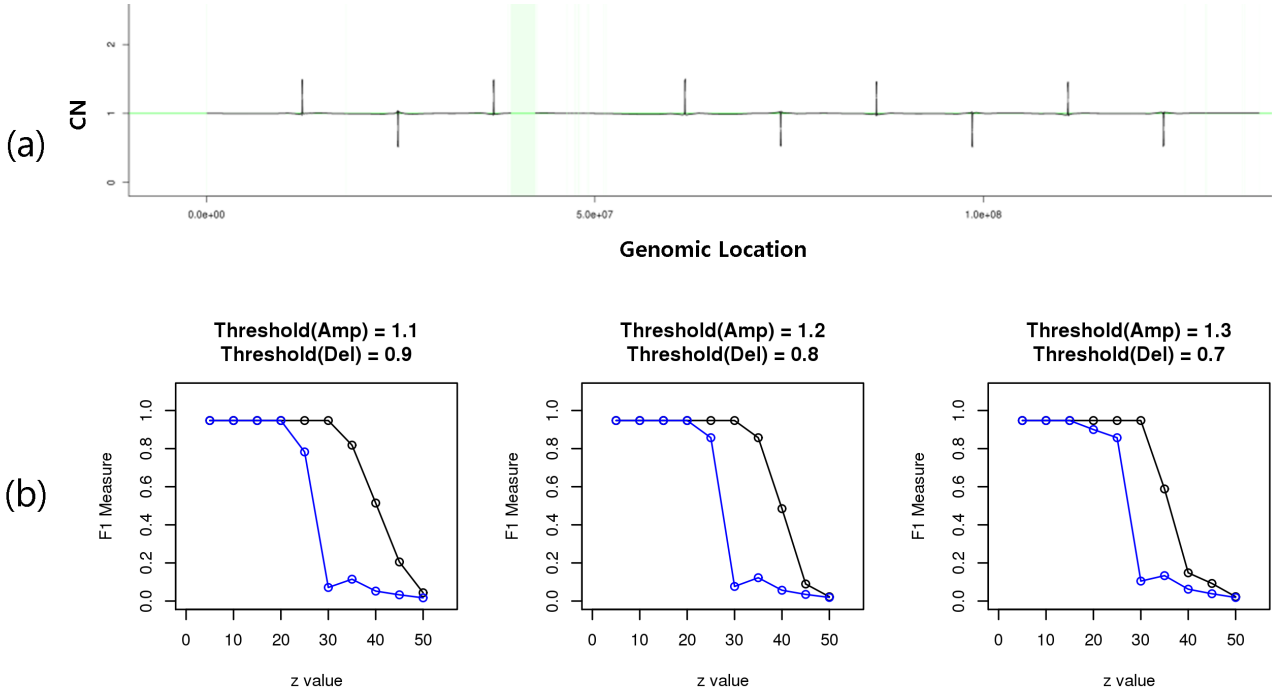

Figure S6. Simulated sequence data with *CN* variation and the F1 measures for detecting these events

2.4 The effect of multiresolution decomposition in real cancer data

Supplementary Figure S7 shows the estimated probability density plots of denoised *CN* ratio windows. Figure S7 (a) is the distribution of CN_{raw}, and Figure S7 (b), (c), and (d) presents the estimated distribution of CN_{Wabico} from the GC bias corrections with decomposition levels 1, 5, and 10. In Figure S7 (a), the peaks representing different *CN*s are not distinguishable due to severe GC bias. As the decomposition levels increase, the peaks become more distinguishable.

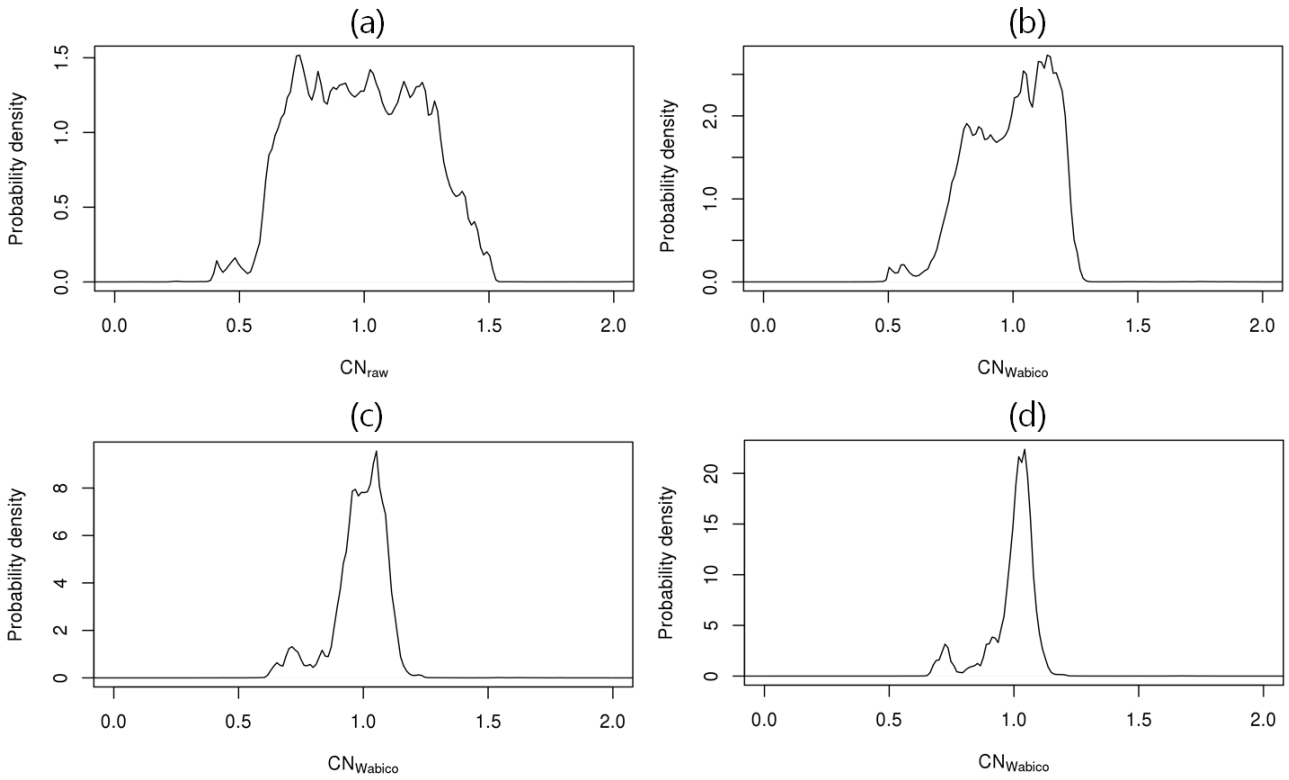

Figure S7. The distribution of *CN* values from TCGA 1444-01A samples

2.5 Standard deviations (SDs) of corrected *CN* signals from paired normal WGS data

We checked the GC bias correction performance using the WGS data of paired normal samples from the 37 patients with GBM. We used SDs as a measure of GC bias correction because DOC_{raw} in the normal sample is largely affected by GC bias but is less affected by structural variations such as *CN* variations. If Wabico controls GC bias better than other methods, the SDs yielded by CN_{Wabico} will be smaller than those from other methods. Table S1 indicates that the SD of CN_{Wabico} is smaller than that of $CN_{BIC-Seq2}$ in most of the paired normal samples from the patients with GBM. This finding suggests that Wabico controls GC bias better than BIC-Seq2 does.

TCGA-ID	SDs CN_{Wabico}	SDs $CN_{BIC-Seq2}$	Smaller SDs
0125-10A	8.41615	8.65231	Wabico
0145-10A	8.73036	9.00338	Wabico
0152-10A	13.14883	17.12831	Wabico
0157-10A	6.11932	6.47396	Wabico
0171-10A	8.15080	8.31309	Wabico
0185-10B	10.87579	13.44248	Wabico
0190-10B	8.63151	8.87745	Wabico
0210-10A	11.68024	12.12270	Wabico
0211-10A	10.84795	11.21337	Wabico
0214-10A	12.17558	12.36649	Wabico
0648-10A	12.13205	12.45086	Wabico
0686-10A	12.00894	12.04671	Wabico
0744-10A	8.83298	8.94882	Wabico
0745-10A	10.33604	10.39009	Wabico
1034-10A	11.08199	11.34486	Wabico
1389-10D	9.89502	10.17476	Wabico
1402-10A	13.65831	13.95338	Wabico
1444-10A	10.33847	10.35900	Wabico
1823-10A	6.22274	6.39560	Wabico
1831-10A	11.67000	11.44439	BIC-Seq2
1970-10A	9.85956	10.16090	Wabico
2483-10A	6.61742	6.85258	Wabico
2485-10A	7.41802	7.47725	Wabico
2523-10A	10.01310	10.18712	Wabico
2528-10A	10.28809	10.48268	Wabico
2554-10A	5.13198	5.49989	Wabico
2557-10A	8.35728	8.47360	Wabico
2570-10A	10.47309	10.86494	Wabico
2620-10A	11.19813	11.31676	Wabico
2624-10A	9.68137	10.09514	Wabico
2629-10A	11.65936	11.86532	Wabico
5132-10A	9.16827	9.21296	Wabico
5135-10A	9.77682	9.77734	Wabico
5411-10A	10.81661	10.76831	BIC-Seq2
5415-10A	11.00828	11.17948	Wabico
5651-10A	5.48416	6.29573	Wabico
5960-10A	11.69929	11.94971	Wabico

Table S1. A comparison of SDs between denoising results of Wabico and of the BIC-Seq2 normalization method for paired normal WGS samples from the patients with GBM

2.6 A comparison of GC correction methods for the identification of cancer-related genes with focal aberrations

In Tables 1, 2, and 3 in the main text, we revealed that Wabico identified most of the pairs of WGS cancer samples and cancer-related genes with focal aberrations for WGS datasets of three cancer types. In this section, we compared data on known cancer genes identified by a segmentation method after GC bias correction by BIC-Seq2–based expected read counts and by Wabico. Here, we ran the BIC-Seq2 segmentation program with $\lambda = 25$ for all samples. The inputs of the BIC-Seq2 segmentation program for Wabico are DOC_{raw} and $DOC_{i,GC}$. The inputs for BIC-Seq2 are DOC_{raw} and $DOC_{i,expected}$. We first set the thresholds for *CN* amplification and deletion to 1.1 and 0.9, respectively. If the *CN* values of a BIC-Seq2 segment were greater than or less than the thresholds of *CN* amplification or deletion, the segment was regarded as *CN* amplification or deletion, respectively.

Tables S2, S3, and S4 present the numbers of genes identified by both methods, not identified by either method, those identified only by Wabico, and those identified only by the BIC-Seq2–based expected read count. In the GBM dataset, two alterations around genes *FGFR3* and *QKI* were identified only by Wabico. In the case of LUAD and OVC datasets, the numbers of genes identified by Wabico and BIC-Seq2 were the identical. This is because the cancer genes for this test were chosen conservatively: those occurring in both WGS and SNP6 datasets. Note that because the BIC-Seq2 segmentation method may not detect some focal aberrant regions or more genes, Wabico results in Supplementary Tables S2, S3, and S4 differ from those in Tables 1, 2, and 3 in the main text, in which a cancer gene was assumed to be identified if the two conditions in Section 2.2 of Supplementary Material were satisfied.

Chr	Name	Type	Total	Identified by both methods	Not identified by either method	Identified only by Wabico	Identified only by BIC-Seq2 expected read count
1	MDM4	amp	6	5	1	0	0
4	FGFR3	amp	4	2	1	1	0
4	PDGFRA	amp	6	6	0	0	0
6	QKI	del	3	2	0	1	0
7	EGFR	amp	23	22	1	0	0
7	CDK6	amp	4	4	0	0	0
9	CDKN2A	del	14	14	0	0	0
9	CDKN2B	del	13	13	0	0	0
10	PTEN	del	3	1	2	0	0
10	FGFR2	amp	2	2	0	0	0
12	CCND2	amp	2	2	0	0	0
12	CDK4	amp	11	10	1	0	0
12	MDM2	amp	7	7	0	0	0
17	GRB2	amp	2	1	1	0	0

Table S2. *CN*-altered GBM-related genes

Chr	Name	Type	Total	Identified by both methods	Not identified by either method	Identified only by Wabico	Identified only by BIC-Seq2 expected read count
5	TERT	amp	2	1	1	0	0
5	PDE4D	del	3	3	0	0	0
9	PTPRD	del	2	1	1	0	0
9	CDKN2A	del	2	1	1	0	0
12	MDM2	amp	2	2	0	0	0
19	CCNE1	amp	2	2	0	0	0

Table S3. *CN*-altered LUAD-related genes

Chr	Name	Type	Total	Identified by both methods	Not identified by either method	Identified only by Wabico	Identified only by BIC-Seq2 expected read count
1	MYCL	amp	7	7	0	0	0
1	MCL1	amp	4	4	0	0	0
3	MECOM	amp	7	7	0	0	0
4	TACC3	amp	5	5	0	0	0
4	ANKRD17	amp	2	2	0	0	0
5	TERT	amp	4	4	0	0	0
6	ID4	amp	4	4	0	0	0
8	SOX17	amp	7	7	0	0	0
8	MYC	amp	16	16	0	0	0
10	PTEN	del	6	6	0	0	0
11	ALG8	amp	6	6	0	0	0
12	KRAS	amp	4	4	0	0	0
13	RB1	del	4	4	0	0	0
14	METTL17	amp	3	3	0	0	0
17	NF1	del	3	3	0	0	0
19	CCNE1	amp	16	16	0	0	0

Table S4. *CN*-altered OVC-related genes

2.7 Correlation with TCGA level 3 segments

We also checked the GC bias correction performance based on the correlation between the *CN* ratio signal from cancer WGS data corrected by Wabico or by the other method and TCGA level 3 segments from SNP6 microarray data. The correlation was calculated from most of the markers of the whole genome. Tables S5, S6, S7 present the results of correlation comparisons. The correlations between denoised CN_{Wabico} and the level 3 segments were higher than the correlations between $CN_{BIC-Seq2}$ and the level 3 segments for all three cancer datasets.

Decomposition level	TCGA-ID	The number of markers	Correlation Wabico	Correlation BIC-Seq2 expected read counts	Higher Correlation
10	0125-01A	22625771	0.74695	0.74616	Wabico
9	0171-01A	22631143	0.84178	0.83887	Wabico
6	0190-01A	22630269	0.79130	0.77644	Wabico
9	0210-01A	22625702	0.90437	0.90074	Wabico
7	0211-01A	22627501	0.70366	0.70345	Wabico
6	0214-01A	22631765	0.90261	0.89866	Wabico
9	0686-01A	22629667	0.75410	0.74854	Wabico
9	0744-01A	22623955	0.62172	0.61785	Wabico
9	0745-01A	22630160	0.74717	0.74528	Wabico
4	1034-01A	22627351	0.91839	0.91388	Wabico
6	1389-01A	22620301	0.70999	0.70971	Wabico
9	1402-01A	22619202	0.87348	0.87339	Wabico
10	1444-01A	22626541	0.72705	0.67039	Wabico
10	1823-01A	22625406	0.81437	0.77784	Wabico
9	1831-01A	22625072	0.76013	0.75957	Wabico
9	1970-01A	22620896	0.86389	0.85905	Wabico
10	2483-01A	22624820	0.86174	0.84393	Wabico
10	2485-01A	22626295	0.92895	0.91557	Wabico
10	2523-01A	22627309	0.74920	0.74473	Wabico
9	2528-01A	22623819	0.90962	0.89178	Wabico
9	2554-01A	22625032	0.83224	0.82550	Wabico
9	2557-01A	22622928	0.86343	0.83415	Wabico
10	2570-01A	22627954	0.88411	0.87794	Wabico
9	2620-01A	22622952	0.90755	0.88536	Wabico
9	2629-01A	22630285	0.91977	0.91659	Wabico
9	5132-01A	22630706	0.65878	0.65802	Wabico
10	5135-01A	22626581	0.89185	0.88995	Wabico
10	5411-01A	22631275	0.93046	0.92763	Wabico
4	5415-01A	22627619	0.79557	0.79448	Wabico
10	5651-01A	22627308	0.92028	0.91787	Wabico
4	0145-01A	22629188	0.72399	0.72693	BIC-Seq2
4	0152-01A	22628657	0.79874	0.81055	BIC-Seq2
9	0157-01A	22631260	0.78187	0.79701	BIC-Seq2
4	0185-01A	22631658	0.81301	0.82461	BIC-Seq2
4	0648-01A	22626194	0.78466	0.79311	BIC-Seq2
9	2624-01A	22628568	0.88774	0.88785	BIC-Seq2
3	5960-01A	22626625	0.82487	0.82526	BIC-Seq2

Table S5. A correlation comparison between denoising results of Wabico and of the BIC-Seq2 normalization method among TCGA patients with GBM

Decomposition level	TCGA-ID	The number of markers	Correlation Wabico	Correlation BIC-Seq2 expected read counts	Higher Correlation
9	1678-01A	22626932	0.93599	0.92184	Wabico
9	1680-01A	22630112	0.80433	0.15984	Wabico
9	2659-01A	22630085	0.94380	0.74516	Wabico
6	4389-01A	22623089	0.88700	0.86593	Wabico
10	4395-01A	22625708	0.96067	0.95929	Wabico
9	4397-01A	22625865	0.94850	0.54057	Wabico
9	4398-01A	22629115	0.91197	0.86729	Wabico
9	4420-01A	22625026	0.90734	0.81480	Wabico
9	4422-01A	22622926	0.91878	0.90164	Wabico
4	4432-01A	19310353	0.86456	0.56788	Wabico
10	5066-01A	22632040	0.57079	0.55961	Wabico
10	5147-01A	22628556	0.91424	0.89713	Wabico
10	5429-01A	22624208	0.79308	0.78263	Wabico
10	6203-01A	22632952	0.49125	0.40877	Wabico
3	6215-01A	22631412	0.89621	0.89368	Wabico
10	6597-01A	22629794	0.92709	0.92684	Wabico
10	6840-01A	22628201	0.94712	0.94668	Wabico
4	7030-01A	22632437	0.21575	0.19900	Wabico
10	7146-01A	22622942	0.95254	0.95228	Wabico
10	7156-01A	22631118	0.87690	0.86745	Wabico
4	7158-01A	22629705	0.93027	0.92891	Wabico
2	7281-01A	22627941	0.81813	0.81805	Wabico
4	8171-01A	22624532	0.93504	0.93355	Wabico
6	8299-01A	22631464	0.30993	0.29978	Wabico
5	4396-01A	22628313	0.93853	0.94005	BIC-Seq2
9	6148-01A	22630998	0.49032	0.49467	BIC-Seq2
10	7143-01A	22632203	0.92474	0.92592	BIC-Seq2
10	7535-01A	22629061	0.92682	0.92727	BIC-Seq2

Table S6. A correlation comparison between denoising results of Wabico and of the BIC-Seq2 normalization method among TCGA patients with LUAD

Decomposition level	TCGA-ID	The number of markers	Correlation Wabico	Correlation BIC-Seq2 expected read counts	Higher Correlation
5	0937-01A	22630047	0.89750	0.88973	Wabico
10	1103-01A	22620527	0.30826	0.29990	Wabico
4	1110-01A	22622390	0.90271	0.89675	Wabico
10	1118-01A	22620989	0.83635	0.79271	Wabico
9	1331-01A	22626256	0.71349	0.68744	Wabico
3	1347-01A	22627157	0.83732	0.83011	Wabico
9	1349-01A	22630118	0.84986	0.82543	Wabico
9	1367-01A	22628864	0.79912	0.77501	Wabico
4	1419-01A	22627522	0.80427	0.80135	Wabico
10	1487-01A	22618665	0.86635	0.86536	Wabico
3	1514-01A	22624418	0.85009	0.84306	Wabico
3	1548-01A	22622391	0.92173	0.91324	Wabico
3	1552-01A	22619514	0.89411	0.88990	Wabico
7	1557-01A	22616063	0.92939	0.92743	Wabico
4	1558-01A	22619862	0.89759	0.89479	Wabico
5	1570-01A	22618494	0.73106	0.72599	Wabico
4	1571-01A	22625573	0.88744	0.88315	Wabico
5	1574-01A	22620796	0.91430	0.90933	Wabico
3	1632-01A	22625070	0.89318	0.89191	Wabico
8	1666-01A	22618236	0.89285	0.87768	Wabico
4	2000-01A	22608363	0.90242	0.89948	Wabico
7	2024-01A	22622395	0.87427	0.86323	Wabico
4	2045-01A	22622572	0.77135	0.76298	Wabico
4	2290-01A	22619634	0.83968	0.82642	Wabico
5	2391-01A	22624850	0.86298	0.85531	Wabico
4	0723-01A	22621857	0.85925	0.87614	BIC-Seq2
7	0727-01A	22624670	0.94080	0.94115	BIC-Seq2
9	0751-01A	22623404	0.77046	0.78002	BIC-Seq2
9	0890-01A	22627744	0.91453	0.91598	BIC-Seq2
9	0906-01A	22626531	0.93535	0.93870	BIC-Seq2
9	0912-01A	22626121	0.92328	0.92462	BIC-Seq2
9	0934-01A	22622161	0.81137	0.81411	BIC-Seq2
9	0938-01A	22627646	0.94031	0.94565	BIC-Seq2
4	0980-01A	22628616	0.86281	0.88106	BIC-Seq2
10	0982-01A	22626987	0.43752	0.44537	BIC-Seq2
6	1124-01A	22618611	0.93246	0.93333	BIC-Seq2
3	1411-01A	22628819	0.82609	0.85178	BIC-Seq2
10	1466-01A	22609694	0.90455	0.90635	BIC-Seq2
8	1477-01A	22629530	0.80168	0.80325	BIC-Seq2
10	1491-01A	22621971	0.92001	0.92346	BIC-Seq2
10	1542-01A	22625299	0.92702	0.93046	BIC-Seq2
9	1544-01A	22620119	0.91634	0.92072	BIC-Seq2
6	1562-01A	22627541	0.88394	0.88476	BIC-Seq2
9	1614-01A	22623535	0.92920	0.93528	BIC-Seq2
10	1634-01A	22623101	0.93316	0.93670	BIC-Seq2
10	2050-01A	22619652	0.87853	0.88186	BIC-Seq2
10	2400-01A	22620683	0.90527	0.90654	BIC-Seq2

Table S7. A correlation comparison between denoising results of Wabico and those of the BIC-Seq2 normalization method among TCGA patients with OVC

2.8 Comparisons of *CN* segments after GC correction

We compared the performance of GC bias correction methods by segmentation results. For *CN* segmentation, we used a BIC-Seq2 segmentation algorithm with $\lambda = 25$ for all the samples. The inputs of the BIC-Seq2 segmentation program for Wabico were DOC_{raw} and $DOC_{i,GC}$. The inputs for BIC-Seq2 were DOC_{raw} and $DOC_{i,expected}$. We compared these segments obtained after application of different GC bias correction methods to the SNP level 3 segment for the same samples from TCGA cancer patients. We determined whether the segment from WGS data is TP, FP, TN, or FN according to the threshold and the overlapping segment from SNP array level 3 data. We set the thresholds for *CN* amplification to 1.1, and *CN* deletion to 0.9. SNP array level 3 segments having *CN* values greater than the amplification threshold were considered true amplification, and those smaller than the deletion threshold were regarded as a true deletion. Similarly, the WGS segments were predicted as amplification or deletion on the basis of the threshold, and these predicted WGS segments were compared with SNP array segments for calculating precision, recall, and F1-scores. Before we applied the above measures to the WGS segments, we made the genomic ranges of segments from Wabico equal to the genomic ranges of segments from BIC-Seq2 by dividing the original segments into subsegments to make the comparison conditions the same. When the segment from Wabico overlapped with the segment of BIC-Seq2, these segments were divided into subsegments. One subsegment consisted of overlapping parts in the original segment, and the other subsegments consisted of nonoverlapping parts in the original segment. Then, because this subsegmentation generated too many small segments, we excluded segments whose size was less than 10,000 base pairs for the comparison so that the performance was not affected by these small segments. Tables S8, S9, and S10 show the results on precision, recall, and F1-scores calculated from TP, FP, TN, and FN. Wabico yielded higher F1 scores for more WGS samples than did the BIC-Seq2–based expected read count for all three cancer types.

TCGA-ID	Num CN Segs	Prec Wabico	Prec BIC2	Prec Better	Racall Wabico	Recall BIC2	Recall Better	F1 Wabico	F1 BIC2	F1 Better
0125-01A	139	0.7692	0.7692	SAME	0.8929	0.8929	SAME	0.8264	0.8264	SAME
0145-01A	169	0.7857	0.8077	BIC-Seq2	0.8919	0.8514	Wabico	0.8354	0.8289	Wabico
0152-01A	94	0.9242	0.9492	BIC-Seq2	0.9104	0.8358	Wabico	0.9173	0.8889	Wabico
0157-01A	88	0.8378	0.8611	BIC-Seq2	0.8857	0.8857	SAME	0.8611	0.8732	BIC-Seq2
0171-01A	152	0.7436	0.7467	BIC-Seq2	0.8657	0.8358	Wabico	0.8000	0.7887	Wabico
0185-01A	93	0.8163	0.8444	BIC-Seq2	0.9302	0.8837	Wabico	0.8696	0.8636	Wabico
0190-01A	156	0.7297	0.7286	Wabico	0.8060	0.7612	Wabico	0.7660	0.7445	Wabico
0210-01A	125	0.7966	0.7966	SAME	0.9038	0.9038	SAME	0.8468	0.8468	SAME
0211-01A	138	0.8933	0.8933	SAME	0.9437	0.9437	SAME	0.9178	0.9178	SAME
0214-01A	130	0.7031	0.7541	BIC-Seq2	0.8654	0.8846	BIC-Seq2	0.7759	0.8142	BIC-Seq2
0648-01A	119	0.7910	0.8125	BIC-Seq2	0.9464	0.9286	Wabico	0.8618	0.8667	BIC-Seq2
0686-01A	187	0.9211	0.8908	Wabico	0.9130	0.9217	BIC-Seq2	0.9170	0.9060	Wabico
0744-01A	139	0.8889	0.8571	Wabico	0.9275	0.8696	Wabico	0.9078	0.8633	Wabico
0745-01A	132	0.7260	0.7324	BIC-Seq2	0.8983	0.8814	Wabico	0.8030	0.8000	Wabico
1034-01A	116	0.8333	0.8305	Wabico	0.8621	0.8448	Wabico	0.8475	0.8376	Wabico
1389-01A	120	0.6275	0.6122	Wabico	0.8000	0.7500	Wabico	0.7033	0.6742	Wabico
1402-01A	106	0.7667	0.7931	BIC-Seq2	0.9583	0.9583	SAME	0.8519	0.8679	BIC-Seq2
1444-01A	172	0.5231	0.6034	BIC-Seq2	0.7391	0.7609	BIC-Seq2	0.6126	0.6731	BIC-Seq2
1823-01A	210	0.9452	0.9527	BIC-Seq2	0.7419	0.8656	BIC-Seq2	0.8313	0.9070	BIC-Seq2
1831-01A	86	0.7353	0.7353	SAME	0.8065	0.8065	SAME	0.7692	0.7692	SAME
1970-01A	99	0.8644	0.8667	BIC-Seq2	0.8644	0.8814	BIC-Seq2	0.8644	0.8739	BIC-Seq2
2483-01A	433	0.6552	0.7074	BIC-Seq2	0.8313	0.8313	SAME	0.7328	0.7644	BIC-Seq2
2485-01A	92	0.9091	0.9268	BIC-Seq2	0.8511	0.8085	Wabico	0.8791	0.8636	Wabico
2523-01A	145	0.7547	0.7414	Wabico	0.8163	0.8776	BIC-Seq2	0.7843	0.8037	BIC-Seq2
2528-01A	147	0.8281	0.7910	Wabico	0.8833	0.8833	SAME	0.8548	0.8346	Wabico
2554-01A	99	0.8958	0.8936	Wabico	0.8958	0.8750	Wabico	0.8958	0.8842	Wabico
2557-01A	113	0.8750	0.7887	Wabico	0.9180	0.9180	SAME	0.8960	0.8485	Wabico
2570-01A	108	0.5526	0.5263	Wabico	0.8750	0.8333	Wabico	0.6774	0.6452	Wabico
2620-01A	108	0.7292	0.7955	BIC-Seq2	0.8333	0.8333	SAME	0.7778	0.8140	BIC-Seq2
2624-01A	101	0.8491	0.8600	BIC-Seq2	0.9184	0.8776	Wabico	0.8824	0.8687	Wabico
2629-01A	111	0.8500	0.8833	BIC-Seq2	0.5862	0.6092	BIC-Seq2	0.6939	0.7211	BIC-Seq2
5132-01A	162	0.8243	0.9077	BIC-Seq2	0.8026	0.7763	Wabico	0.8133	0.8369	BIC-Seq2
5135-01A	194	0.8222	0.8427	BIC-Seq2	0.9367	0.9494	BIC-Seq2	0.8757	0.8929	BIC-Seq2
5411-01A	186	0.6029	0.5821	Wabico	0.9111	0.8667	Wabico	0.7257	0.6964	Wabico
5415-01A	164	0.8049	0.8125	BIC-Seq2	0.9041	0.8904	Wabico	0.8516	0.8497	Wabico
5651-01A	203	0.8161	0.7640	Wabico	0.8068	0.7727	Wabico	0.8114	0.7684	Wabico
5960-01A	139	0.9639	0.9639	SAME	0.6504	0.6504	SAME	0.7767	0.7767	SAME

Table S8. A comparison on precision, recall, and F1-scores of segments from Wabico and BIC-Seq2 for TCGA patients with GBM

TCGA-ID	Num CN Segs	Prec Wabico	Prec BIC2	Prec Better	Racall Wabico	Recall BIC2	Recall Better	F1 Wabico	F1 BIC2	F1 Better
1678-01A	129	0.9494	0.9481	Wabico	0.9259	0.9012	Wabico	0.9375	0.9241	Wabico
1680-01A	64	0.8056	0.7667	Wabico	0.8788	0.6970	Wabico	0.8406	0.7302	Wabico
2659-01A	82	0.9070	0.8824	Wabico	0.7091	0.8182	BIC-Seq2	0.7959	0.8491	BIC-Seq2
4389-01A	145	0.8750	0.8621	Wabico	0.7549	0.7353	Wabico	0.8105	0.7937	Wabico
4395-01A	165	0.8952	0.8785	Wabico	0.8468	0.8468	SAME	0.8704	0.8624	Wabico
4396-01A	170	0.9785	0.9783	Wabico	0.6947	0.6870	Wabico	0.8125	0.8072	Wabico
4397-01A	197	0.7812	0.8065	BIC-Seq2	0.8621	0.8621	SAME	0.8197	0.8333	BIC-Seq2
4398-01A	227	0.8284	0.8952	BIC-Seq2	0.8740	0.8740	SAME	0.8506	0.8845	BIC-Seq2
4420-01A	304	0.7250	0.7559	BIC-Seq2	0.7178	0.7970	BIC-Seq2	0.7214	0.7759	BIC-Seq2
4422-01A	189	0.8762	0.8559	Wabico	0.9200	0.9500	BIC-Seq2	0.8976	0.9005	BIC-Seq2
4432-01A	149	0.9670	0.9438	Wabico	0.8224	0.7850	Wabico	0.8889	0.8571	Wabico
5066-01A	637	0.2953	0.2627	Wabico	0.7732	0.6907	Wabico	0.4274	0.3807	Wabico
5147-01A	486	0.8340	0.7584	Wabico	0.8340	0.7698	Wabico	0.8340	0.7640	Wabico
5429-01A	148	0.7097	0.7667	BIC-Seq2	0.8148	0.8519	BIC-Seq2	0.7586	0.8070	BIC-Seq2
6148-01A	67	0.4706	0.4706	SAME	1.0000	1.0000	SAME	0.6400	0.6400	SAME
6203-01A	1982	0.3189	0.2909	Wabico	0.4591	0.4715	BIC-Seq2	0.3764	0.3598	Wabico
6215-01A	108	0.9348	0.9348	SAME	0.8958	0.8958	SAME	0.9149	0.9149	SAME
6597-01A	196	0.8496	0.8485	Wabico	0.8760	0.8682	Wabico	0.8626	0.8582	Wabico
6840-01A	225	0.9107	0.9107	SAME	0.8453	0.8453	SAME	0.8768	0.8768	SAME
7030-01A	77	0.2857	0.3636	BIC-Seq2	0.5714	0.5714	SAME	0.3810	0.4444	BIC-Seq2
7143-01A	148	0.6753	0.7183	BIC-Seq2	0.8387	0.8226	Wabico	0.7482	0.7669	BIC-Seq2
7146-01A	274	0.9840	0.9842	BIC-Seq2	0.7541	0.7664	BIC-Seq2	0.8538	0.8618	BIC-Seq2
7156-01A	172	0.6197	0.5946	Wabico	0.8800	0.8800	SAME	0.7273	0.7097	Wabico
7158-01A	81	0.7692	0.8333	BIC-Seq2	0.9524	0.9524	SAME	0.8511	0.8889	BIC-Seq2
7281-01A	79	0.8261	0.8372	BIC-Seq2	0.8636	0.8182	Wabico	0.8444	0.8276	Wabico
7535-01A	109	0.9294	0.9286	Wabico	0.8587	0.8478	Wabico	0.8927	0.8864	Wabico
8171-01A	257	0.9837	0.9836	Wabico	0.9731	0.9677	Wabico	0.9784	0.9756	Wabico
8299-01A	95	0.2308	0.2308	SAME	1.0000	1.0000	SAME	0.3750	0.3750	SAME

Table S9. A comparison on precision, recall, and F1-scores of segments from Wabico and BIC-Seq2 among TCGA patients with LUAD

TCGA-ID	Num CN Segs	Prec Wabico	Prec BIC2	Prec Better	Racall Wabico	Recall BIC2	Recall Better	F1 Wabico	F1 BIC2	F1 Better
0723-01A	978	0.9439	0.9457	BIC-Seq2	0.8913	0.8970	BIC-Seq2	0.9169	0.9207	BIC-Seq2
0727-01A	940	0.8619	0.8673	BIC-Seq2	0.9402	0.9459	BIC-Seq2	0.8994	0.9049	BIC-Seq2
0751-01A	2467	0.9866	0.9774	Wabico	0.7858	0.7939	BIC-Seq2	0.8749	0.8762	BIC-Seq2
0890-01A	1577	0.9513	0.9574	BIC-Seq2	0.6059	0.6120	BIC-Seq2	0.7403	0.7467	BIC-Seq2
0906-01A	1173	0.9773	0.9787	BIC-Seq2	0.5816	0.5807	Wabico	0.7292	0.7289	Wabico
0912-01A	1025	0.9937	0.9917	Wabico	0.5113	0.5113	SAME	0.6752	0.6747	Wabico
0934-01A	1148	0.8836	0.8832	Wabico	0.9705	0.9672	Wabico	0.9250	0.9233	Wabico
0937-01A	1048	0.9759	0.9759	Wabico	0.6969	0.6945	Wabico	0.8132	0.8115	Wabico
0938-01A	1108	0.9611	0.9646	BIC-Seq2	0.7822	0.7802	Wabico	0.8625	0.8626	BIC-Seq2
0980-01A	110	0.8136	0.8103	Wabico	0.8421	0.8246	Wabico	0.8276	0.8174	Wabico
0982-01A	2661	0.9709	0.9773	BIC-Seq2	0.9025	0.9114	BIC-Seq2	0.9354	0.9432	BIC-Seq2
1103-01A	5002	0.9616	0.9551	Wabico	0.6521	0.6680	BIC-Seq2	0.7772	0.7862	BIC-Seq2
1110-01A	1035	0.9305	0.9400	BIC-Seq2	0.9535	0.9564	BIC-Seq2	0.9419	0.9481	BIC-Seq2
1118-01A	1395	0.9112	0.9191	BIC-Seq2	0.7011	0.7002	Wabico	0.7925	0.7948	BIC-Seq2
1124-01A	1156	0.8691	0.8714	BIC-Seq2	0.9122	0.9077	Wabico	0.8901	0.8891	Wabico
1331-01A	1435	0.9368	0.9347	Wabico	0.6953	0.6847	Wabico	0.7982	0.7904	Wabico
1347-01A	1175	0.9042	0.9126	BIC-Seq2	0.8822	0.8981	BIC-Seq2	0.8931	0.9053	BIC-Seq2
1349-01A	1126	0.9060	0.9046	Wabico	0.7308	0.7393	BIC-Seq2	0.8090	0.8136	BIC-Seq2
1367-01A	1167	0.9770	0.9768	Wabico	0.5540	0.5497	Wabico	0.7071	0.7035	Wabico
1411-01A	124	0.8889	0.8919	BIC-Seq2	0.8571	0.7857	Wabico	0.8727	0.8354	Wabico
1419-01A	1219	0.9510	0.9510	SAME	0.5980	0.5980	SAME	0.7343	0.7343	SAME
1466-01A	1901	0.8818	0.8824	BIC-Seq2	0.9057	0.9057	SAME	0.8936	0.8939	BIC-Seq2
1477-01A	951	0.8821	0.8785	Wabico	0.9234	0.9234	SAME	0.9023	0.9004	Wabico
1487-01A	1509	0.9914	0.9922	BIC-Seq2	0.7059	0.6943	Wabico	0.8247	0.8170	Wabico
1491-01A	1237	0.9934	0.9932	Wabico	0.6101	0.5906	Wabico	0.7559	0.7407	Wabico
1514-01A	1026	0.9336	0.9334	Wabico	0.7108	0.7086	Wabico	0.8071	0.8056	Wabico
1542-01A	1362	0.9843	0.9861	BIC-Seq2	0.9700	0.9700	SAME	0.9771	0.9780	BIC-Seq2
1544-01A	1385	0.9573	0.9681	BIC-Seq2	0.9671	0.9661	Wabico	0.9622	0.9671	BIC-Seq2
1548-01A	864	0.9908	0.9908	BIC-Seq2	0.7948	0.7960	BIC-Seq2	0.8820	0.8828	BIC-Seq2
1552-01A	1091	0.9902	0.9916	BIC-Seq2	0.6711	0.6721	BIC-Seq2	0.8000	0.8011	BIC-Seq2
1557-01A	1360	0.8762	0.8751	Wabico	0.8873	0.8853	Wabico	0.8817	0.8802	Wabico
1558-01A	988	0.8478	0.8489	BIC-Seq2	0.8935	0.8935	SAME	0.8700	0.8706	BIC-Seq2
1562-01A	1166	0.9906	0.9925	BIC-Seq2	0.4626	0.4634	BIC-Seq2	0.6306	0.6318	BIC-Seq2
1570-01A	1319	0.9729	0.9729	Wabico	0.9473	0.9459	Wabico	0.9599	0.9592	Wabico
1571-01A	1011	0.9377	0.9445	BIC-Seq2	0.9543	0.9543	SAME	0.9459	0.9494	BIC-Seq2
1574-01A	903	0.9686	0.9705	BIC-Seq2	0.6168	0.6155	Wabico	0.7537	0.7533	Wabico
1614-01A	1111	0.8922	0.8927	BIC-Seq2	0.6936	0.6966	BIC-Seq2	0.7805	0.7826	BIC-Seq2
1632-01A	1177	0.7900	0.7911	BIC-Seq2	0.7016	0.6984	Wabico	0.7431	0.7419	Wabico
1634-01A	909	0.9404	0.9446	BIC-Seq2	0.9485	0.9514	BIC-Seq2	0.9444	0.9480	BIC-Seq2
1666-01A	944	0.9860	0.9860	Wabico	0.5441	0.5419	Wabico	0.7012	0.6994	Wabico
2000-01A	1489	0.9694	0.9716	BIC-Seq2	0.9728	0.9693	Wabico	0.9711	0.9704	Wabico
2024-01A	882	1.0000	1.0000	SAME	0.4846	0.4732	Wabico	0.6528	0.6424	Wabico
2045-01A	1229	0.9579	0.9562	Wabico	0.9838	0.9838	SAME	0.9707	0.9698	Wabico
2050-01A	1293	0.9923	0.9935	BIC-Seq2	0.6090	0.6043	Wabico	0.7548	0.7515	Wabico
2290-01A	1013	0.9103	0.9222	BIC-Seq2	0.6166	0.6287	BIC-Seq2	0.7352	0.7477	BIC-Seq2
2391-01A	1028	0.9598	0.9572	Wabico	0.5196	0.5109	Wabico	0.6742	0.6662	Wabico
2400-01A	1098	0.9260	0.9270	BIC-Seq2	0.9249	0.9237	Wabico	0.9255	0.9254	Wabico

Table S10. A comparison on precision, recall, and F1-scores of segments from Wabico and BIC-Seq2 among TCGA patients with OVC

2.9 Correlation with TCGA level3 segments (2)

Additionally, we compared the performance of Wabico and the performance of the GC correction method used in the CNVkit method. Although CNVkit is mainly focused on the detection of CN variations in WES data, it also supports functions for WGS data processing. We applied the following command to the WGS tumor bam files. "cnvkit.py batch TUMORBAM -r REFERENCE", where TUMORBAM is the location of the WGS bam file and REFERENCE is the location of human reference sequence of hg19. It produces a cnr file. We denote the GC bias corrected copy number ratio signal in the cnr file as CN_{CNVKIT}. This signal consists of 536,913 markers for 22 somatic chromosomes. We call the markers $MARKER_{CNVKIT}$. The median of genomic ranges of $MARKER_{CNVKIT}$ is 5,000 bps. In the previous section, we used 22,641,222 markers that cover 100 uniquely mappable positions for the comparison of CN_{Wabico} and $CN_{BIC-Seq2}$. We call these markers $MARKER_{WABICO}$. The median of genomic ranges of $MARKER_{WABICO}$ is 100 bps. To compare the GC correction result of Wabico to that of CN_{CNVKIT}, we created CN'_{Wabico} from CN_{Wabico}. CN'_{Wabico} consists of 536,913 markers whose genomic ranges are identical to $MARKER_{CNVKIT}$. We first calculated the overlaps of genomic ranges between $MARKER_{WABICO}$ and $MARKER_{CNVKIT}$. Then, we calculated copy number ratios of markers in $MARKER_{CNVKIT}$ by averaging the copy number ratios of $MARKER_{WABICO}$ whose genomic ranges are within the marker. After converting $MARKER_{WABICO}$ to $MARKER_{CNVKIT}$, we calculated the correlation coefficient between denoised CN'_{Wabico} signal and TCGA level 3 segments. We also calculated the correlation coefficient between denoised CN_{CNVKIT} signal and TCGA level 3 segments. Finally, we compared the correlation coefficients from Wabico and CNVkit. Tables S11, S12, and S13 show comparison of correlation coefficients for GBM, LUAD, and OVC datasets. Except for one OV sample (0934-01A), correlation coefficients of Wabico are higher than those of CNVkit for all three tumor datasets. Figure S8 shows an example of the denoised copy number ratios of chromosome 1 (GBM 0745-01A). Note that because paired normal control was not used in Wabico, it was not used in CNVkit as well.

TCGA-ID	The number of markers	Correlation Wabico	Correlation CNVkit (cnr file)	Higher Correlation
0125-01A	535777	0.87848	0.72906	Wabico
0145-01A	535371	0.84404	0.55203	Wabico
0152-01A	535506	0.80275	0.54914	Wabico
0157-01A	535253	0.85851	0.5659	Wabico
0171-01A	535250	0.87298	0.68017	Wabico
0185-01A	535432	0.85779	0.65482	Wabico
0190-01A	535348	0.79844	0.50577	Wabico
0210-01A	535617	0.81826	0.40524	Wabico
0211-01A	535148	0.83815	0.56893	Wabico
0214-01A	535233	0.85735	0.65871	Wabico
0648-01A	535271	0.88541	0.63432	Wabico
0686-01A	534896	0.87435	0.5907	Wabico
0744-01A	535105	0.9109	0.5281	Wabico
0745-01A	535724	0.92286	0.58066	Wabico
1034-01A	535340	0.90172	0.70563	Wabico
1389-01A	535861	0.68329	0.42393	Wabico
1402-01A	535457	0.84944	0.74271	Wabico
1444-01A	535268	0.77853	0.27538	Wabico
1823-01A	535287	0.89335	0.54405	Wabico
1831-01A	535455	0.88079	0.65506	Wabico
1970-01A	535583	0.91541	0.71024	Wabico
2483-01A	534982	0.91899	0.52436	Wabico
2485-01A	535328	0.91614	0.81064	Wabico
2523-01A	535423	0.89347	0.71995	Wabico
2528-01A	534975	0.78696	0.50016	Wabico
2554-01A	535276	0.86637	0.66133	Wabico
2557-01A	535404	0.87807	0.53493	Wabico
2570-01A	535297	0.69657	0.52162	Wabico
2620-01A	535405	0.80463	0.3678	Wabico
2624-01A	535496	0.78823	0.53824	Wabico
2629-01A	535131	0.92174	0.61723	Wabico
5132-01A	535156	0.9104	0.44811	Wabico
5135-01A	535396	0.88525	0.75889	Wabico
5411-01A	535398	0.81256	0.64685	Wabico
5415-01A	535789	0.8991	0.75758	Wabico
5651-01A	535390	0.8919	0.76852	Wabico
5960-01A	535478	0.8917	0.60256	Wabico

Table S11. Performance comparison between denosing results of Wabico and the CNVkit bias correction method for TCGA GBM data.

TCGA-ID	The number of markers	Correlation Wabico	Correlation CNVkit (cnr file)	Higher Correlation
1678-01A	535436	0.91748	0.76154	Wabico
1680-01A	535263	0.87701	0.27079	Wabico
2659-01A	535006	0.94504	0.72945	Wabico
4389-01A	535387	0.89948	0.73512	Wabico
4395-01A	535027	0.94717	0.80927	Wabico
4396-01A	535217	0.91294	0.76112	Wabico
4397-01A	535376	0.94253	0.72946	Wabico
4398-01A	535074	0.91699	0.48248	Wabico
4420-01A	535371	0.93229	0.52854	Wabico
4422-01A	535331	0.91693	0.51375	Wabico
4432-01A	535199	0.90999	0.62904	Wabico
5066-01A	534931	0.65904	0.4766	Wabico
5147-01A	535129	0.9329	0.82013	Wabico
5429-01A	535914	0.78766	0.54683	Wabico
6148-01A	535334	0.53176	0.28568	Wabico
6203-01A	535216	0.59415	0.22676	Wabico
6215-01A	535432	0.89207	0.79872	Wabico
6597-01A	535471	0.91748	0.84928	Wabico
6840-01A	535378	0.94359	0.90599	Wabico
7030-01A	535325	0.22566	0.10859	Wabico
7143-01A	535467	0.9018	0.7823	Wabico
7146-01A	535474	0.92389	0.82847	Wabico
7156-01A	535326	0.87017	0.80056	Wabico
7158-01A	535428	0.93116	0.85631	Wabico
7281-01A	535311	0.83593	0.49722	Wabico
7535-01A	535385	0.93099	0.65685	Wabico
8171-01A	535314	0.92857	0.75315	Wabico
8299-01A	535345	0.28905	0.26113	Wabico

Table S12. Performance comparison between denosing results of Wabico and the CNVkit bias correction method for TCGA LUAD cancer.

TCGA-ID	The number of markers	Correlation Wabico	Correlation CNVkit (cnr file)	Higher Correlation
0723-01A	535311	0.8685	0.82092	Wabico
0727-01A	535303	0.96381	0.88269	Wabico
0751-01A	535023	0.95437	0.75566	Wabico
0890-01A	534848	0.94458	0.63089	Wabico
0906-01A	534889	0.96859	0.77125	Wabico
0912-01A	535703	0.94346	0.75702	Wabico
0934-01A	535139	0.90535	0.9191	CNVkit
0937-01A	535156	0.96038	0.90836	Wabico
0938-01A	535888	0.94128	0.89863	Wabico
0980-01A	535092	0.93491	0.80239	Wabico
0982-01A	535222	0.82582	0.82582	Wabico
1103-01A	535014	0.83316	0.7563	Wabico
1110-01A	535200	0.95996	0.86785	Wabico
1118-01A	535340	0.92693	0.82188	Wabico
1124-01A	534856	0.95779	0.8929	Wabico
1331-01A	535238	0.90827	0.75609	Wabico
1347-01A	535109	0.94368	0.82881	Wabico
1349-01A	534795	0.95574	0.85959	Wabico
1367-01A	534781	0.93368	0.72583	Wabico
1411-01A	535047	0.90668	0.77213	Wabico
1419-01A	535293	0.88039	0.7863	Wabico
1466-01A	535455	0.92799	0.78341	Wabico
1477-01A	535203	0.90315	0.73165	Wabico
1487-01A	534975	0.93157	0.76821	Wabico
1491-01A	535233	0.94436	0.62768	Wabico
1514-01A	535310	0.89836	0.83848	Wabico
1542-01A	535123	0.96828	0.80482	Wabico
1544-01A	535340	0.94092	0.79784	Wabico
1548-01A	535405	0.95661	0.82027	Wabico
1552-01A	535204	0.91993	0.78024	Wabico
1557-01A	535402	0.96169	0.91814	Wabico
1558-01A	535389	0.90476	0.85474	Wabico
1562-01A	535208	0.9433	0.75905	Wabico
1570-01A	535234	0.90889	0.90084	Wabico
1571-01A	535261	0.91429	0.87236	Wabico
1574-01A	535252	0.9586	0.81514	Wabico
1614-01A	535287	0.96064	0.82556	Wabico
1632-01A	535111	0.93261	0.74177	Wabico
1634-01A	535065	0.94719	0.66543	Wabico
1666-01A	535105	0.94086	0.79522	Wabico
2000-01A	535403	0.92733	0.91057	Wabico
2024-01A	535045	0.93471	0.73247	Wabico
2045-01A	535288	0.92253	0.74565	Wabico
2050-01A	535122	0.94872	0.71532	Wabico
2290-01A	534889	0.94096	0.81036	Wabico
2391-01A	535070	0.94336	0.88571	Wabico
2400-01A	535356	0.91264	0.86581	Wabico

Table S13. Performance comparison between denosing results of Wabico and the CNVkit bias correction method for TCGA OVC data.

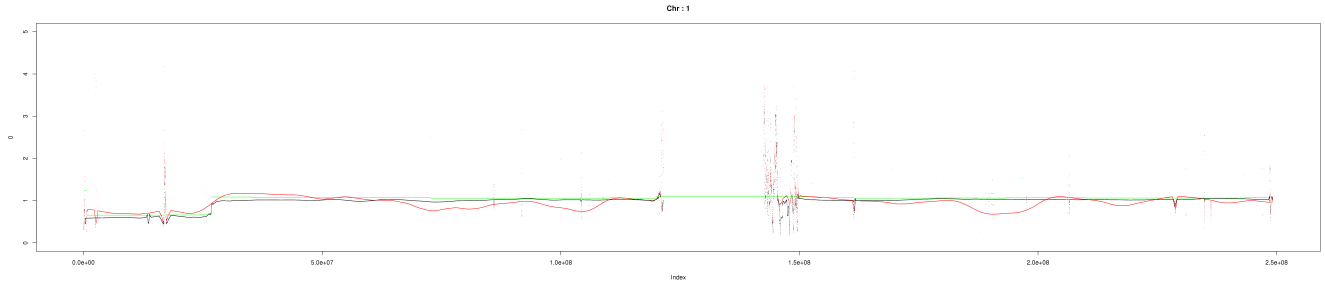

Figure S8. Copy number (CN) Ratio signals of chromosome 1 in TCGA GBM 0745-01A samples. Green lines, red lines, and black lines show TCGA level 3 segments, CN_{CNVKIT} signal, and CN_{Wabico} signal, respectively.

References

1. Jang, H., Hur, Y. & Lee, H. Identification of cancer-driver genes in focal genomic alterations from whole genome sequencing data. Scientific reports **6** (2016).