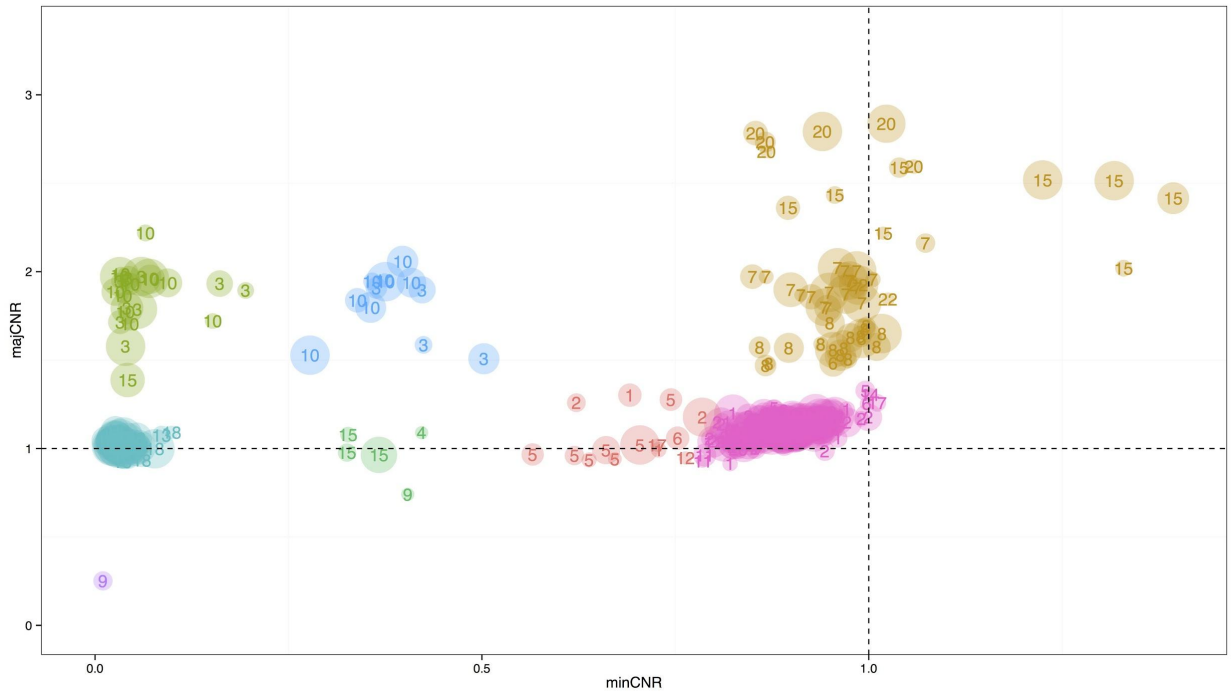
Supplementary figures and tables

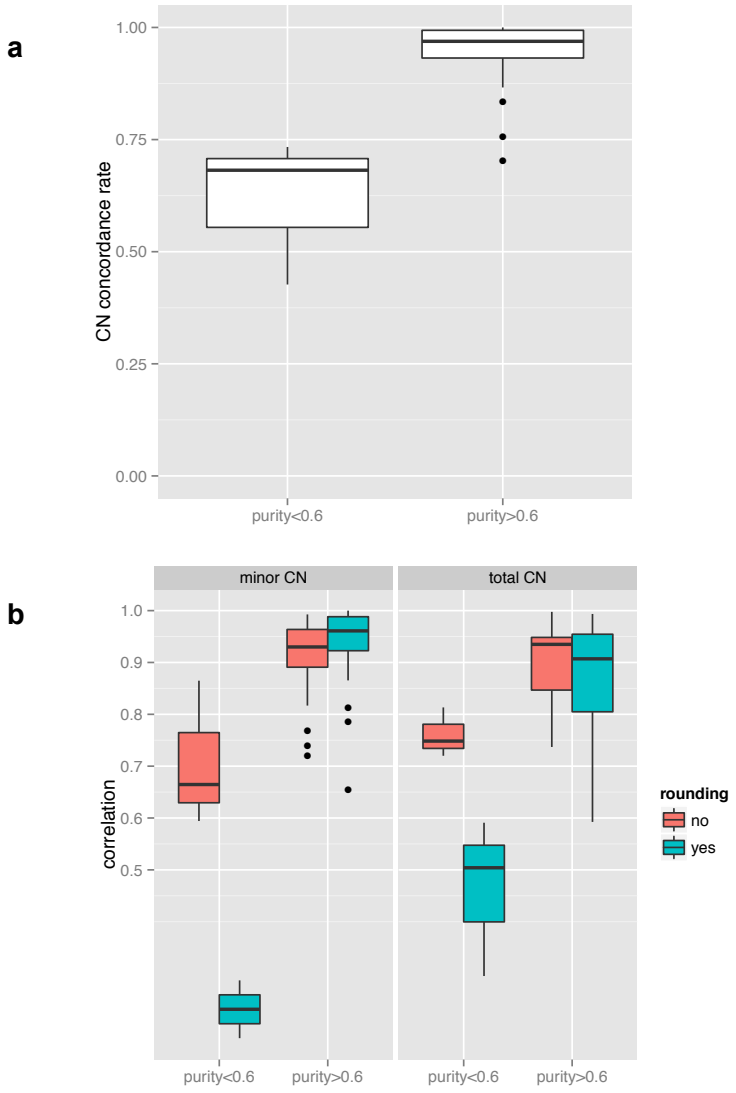
Global copy number profiling of cancer genomes

Wang et al.



Supplementary Figure 1

Assessment of copy number status of one sample based on distance based modified Chinese restaurant process (CRP). The relative ASCN (allele specific copy number) estimates can be calculated based on LRR and BAF. The center of each circle in the plot depicts the minCN and majCN ratios (or relative ASCN) estimated from each genomic segment. The size of the circle indicates the length of one segment. The colors of the circles indicate the grouping results from the Chinese restaurant process (CRP). As shown in the figure, CRP enables the partition of the genome-wide CNA profile into blocks corresponding to different CNA status. The main advantage of the algorithm is that it allows unknown number of clusters without need to model number of clusters directly. The cluster (pink cluster) that is closest to the baseline point (minCN=1 and majCN=1) corresponds to the normal regions, and other cluster status can be inferred accordingly, as described in the supplementary method.

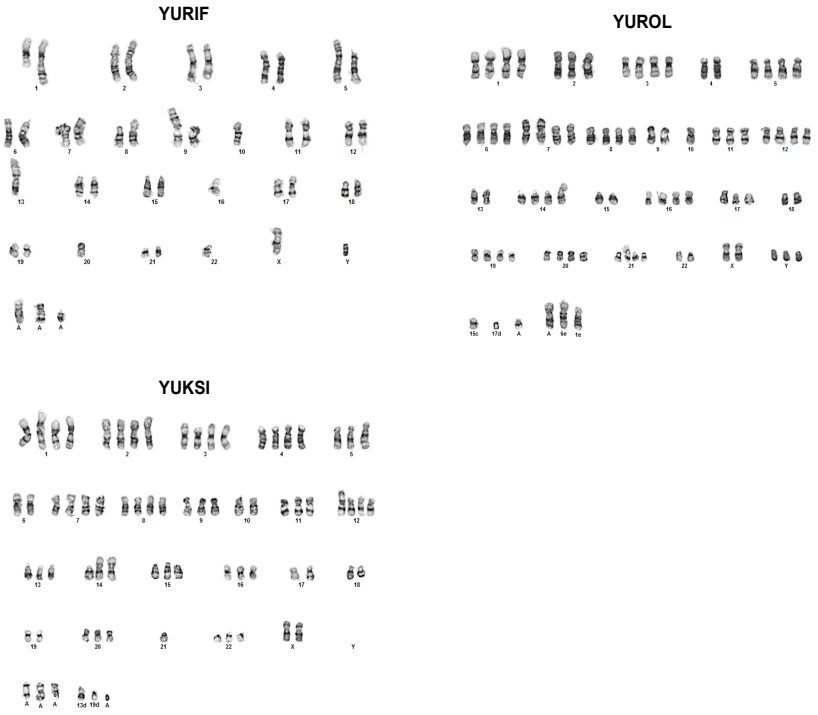


Supplementary Figure 2

Comparison of results from model free and model based approaches (a) and between CLOSE and falcon (b) based on 30 randomly selected samples (sample data summarized in the supplementary table).

(a) The comparison shows a high concordance rate (the median *>95%* when purity>0.6) in segmental copy number calls (gain, loss, normal and CN neutral LOH) generated from model-based (likelihood) and model-free (CRP) approaches implemented in CLOSE. **(b)** Pearson correlation between copy number estimates (minor allele CN and total CN) from falcon (based on read depth models) and CLOSE (based on segmental models) shows high consistency (median >0.9 if purity >0.6) between the analyses from two sources. As shown by the blue boxplots, the correlation reduces significantly when rounding the continuous CN estimates (falcon) to integer if sample purity is low (or when the ploidy is not 2), because the relative copy number will not be equal to the absolute copy number.

a



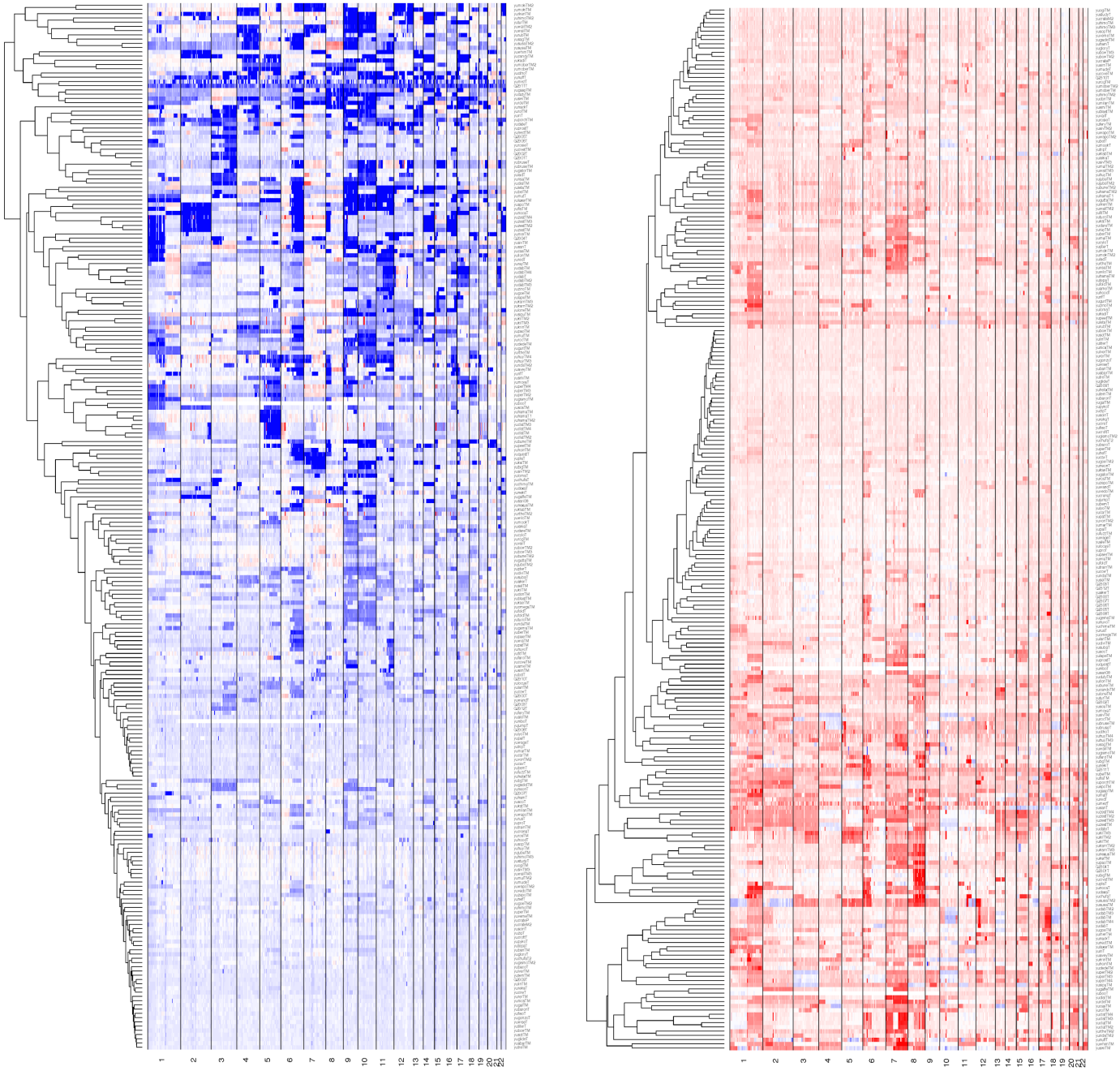
b

Sample ID	Cytogenetic ploidy	Sequenza		CLOSE	
		purity	ploidy	purity	ploidy
YURIF	2n	0.94	4.1	0.96	2
YUROL	4n	0.94	4.4	0.92	4
YUKSI	4n	0.8	6.3	0.82	4
YUSIV	3n	0.43	3.1	0.42	3

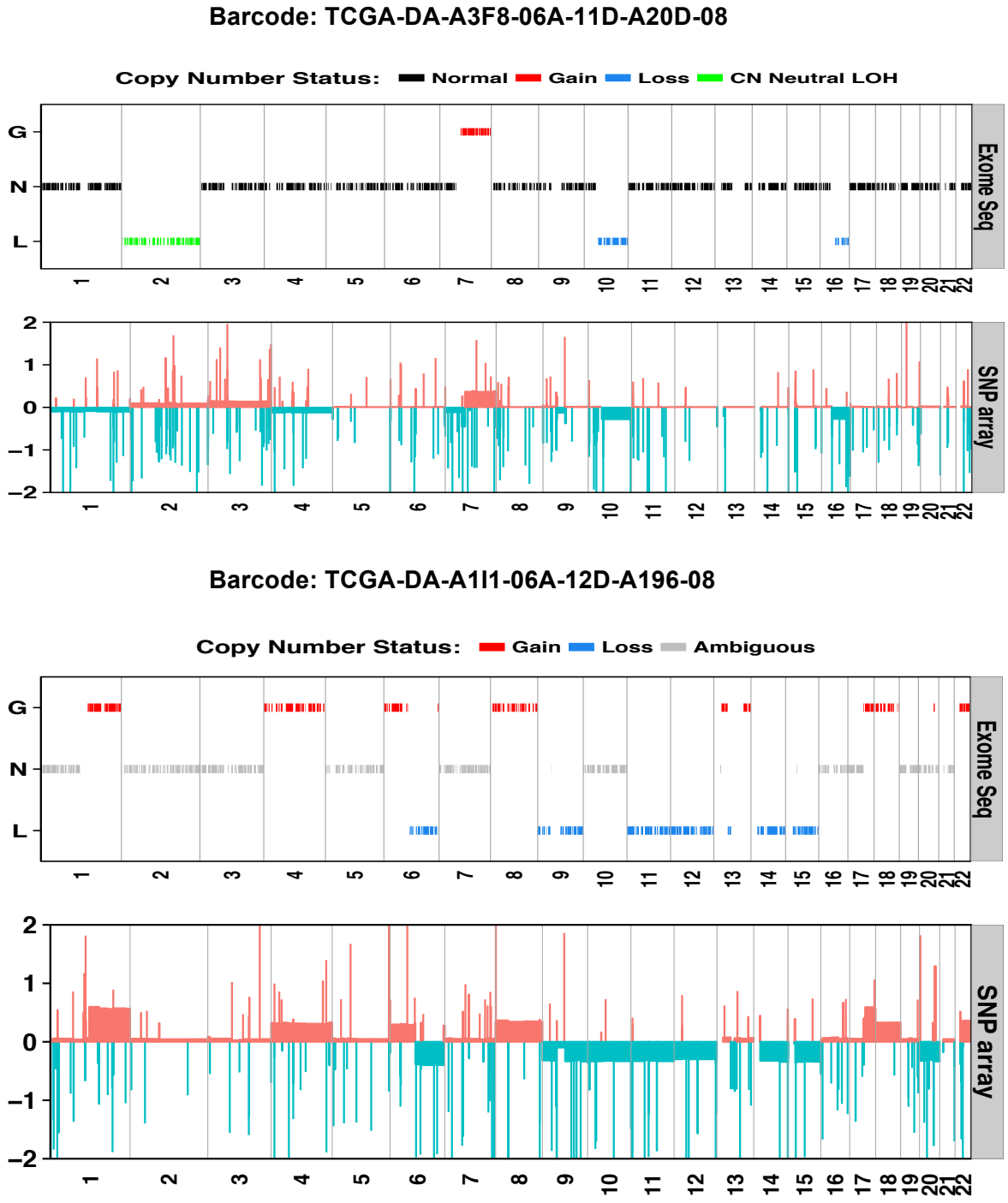
Supplementary Figure 3

Cytogenetic analysis and validation of global ploidy estimation.

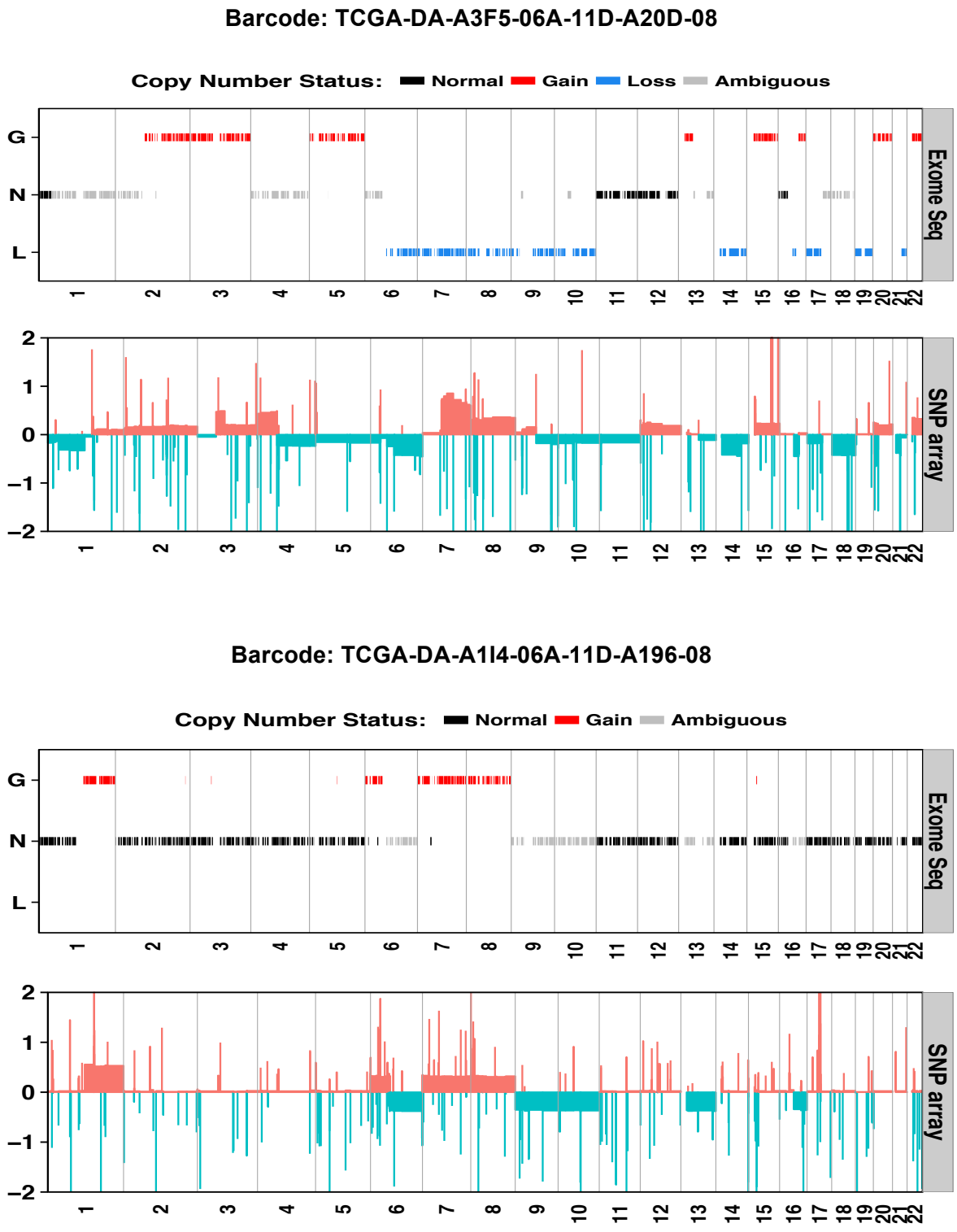
(**a**) Karotype analysis of cell populations from three samples shows that the (average) ploidy of these samples are diploid (YURIF), tetraploid(YUROL) and tetraploid(YUKSI), respectively. (**b**) Comparison of global purity and ploidy estimation from Sequenza and CLOSE. Results show that our approach potentially yields better estimation for the average ploidy level while two methods give consistent estimation on tumor purity. Purity estimates of 30 randomly selected samples (diploid) are listed in the supplementary table 1.



Supplementary Figure 4.
Bivariate hierarchical clustering of 253 melonoma samples based on both minor allele copy number (left panel) and major allele copy number (right panel) estimates. For each sample, we calculated the average minor allele copy number and major allele copy number for each 1Mb genomic window, respectively. Then the averaged copy numbers were log2 transformed. The hierarchical clustering was performed using the Euclidean distance with complete linkage.



Supplementary Figure 5a.
Comparison of global copy number calls from WES and SNP array (Affeymetrix SNP6). The segmented copy number estimates (Segment_mean) based on SNP6 are downloaded from broad firehose website (http://gdac.broadinstitute.org/runs/stddata__2015_04_02).



Supplementary Figure 5b.
Comparison of global copy number calls from WES and SNP array (Affeymetrix SNP6). The segmented copy number estimates (Segment_mean) based on SNP6 are downloaded from broad firehose website (http://gdac.broadinstitute.org/runs/stddata__2015_04_02).

Supplementary Table 1

Tumor samples used in the method comparison. There samples (diploid) are randomly selected from a pool of samples that meet two criteria: 1) estimated purity>0.2 and 2) have at least 2 CNA clusters. The purity estimates in CLOSE are obtained from the canonical point approach as described in the supplementary methods.

Tumor Sample ID	Purity estimates	
	sequenza	CLOSE
yuameTM	0.98	0.95
yuaveyTM	0.97	0.99
yuberTM	0.69	0.65
yubigTM	0.78	0.74
yucandyTM	0.8	0.84
yucivetTM	0.92	0.93
yuclatTM3	0.82	0.83
yudabTM	0.73	0.75
yudedeTM	0.99	0.98
yudexaT	0.82	0.79
yufarciTM	0.63	0.61
yugaffeTM	0.76	0.51
yuhamT	0.37	0.38
yukatTM	0.39	0.43
yukilTM2	0.35	0.33
yuladT	0.67	0.68
yulapeTM	0.9	0.81
yulomaT	0.91	0.91
yumerTM	0.8	0.8
yunackT	0.86	0.89
yunexusTM	0.79	0.77
yuplaT	0.98	0.97
yuprostT	0.74	0.72
yuridaTM2	0.66	0.61
yurifT	0.96	0.94
yurisaTM	0.67	0.68
yurosTM	0.66	0.62
yuwhimTM	0.95	0.99
yuzestTM4	0.93	0.97
yuzinoTM	0.91	0.93