Supplementary Data 1

Performance of CONSERTING in TCGA GBM data

The overall consistency between WGS CNA and SNP array which is measured by F_1 score is slightly lower in TCGA GBM data than that of T-ALL. Several factors contributed to this. First, several cases (e.g. 06-0190-02A) show similar WGS CNA profiles between CONSERTING and BIC-seq, which differ from the corresponding SNP array profiles. This suggests the differences in the tumor specimen used for SNP array and WGS are likely an important contributor to cross-platform inconsistency of CNA analysis. Second, no manual curation was performed for the CNAs derived from the SNP array data for the GBM data set while those for the T-ALL were manually curated. Third, the WGS quality of TCGA GBM data is lower than that of T-ALL (Supplementary Fig. 9).

Supplementary Data 2

CONSERTING for enhancing chromothripsis analysis in SJRB003

The purity of retinoblastoma tumor SJRB003 was estimated to be 95%, and the absence of wild-type *RB1* gene in this tumor supports this high level[1]. CONSERTING detected 12 amplification CNAs scattered across a 78 Mb region on chromosome 1q (chr1: 164 – 242 Mb), 11 of which were estimated to have 0.5 fold amplification, and only 5 were found by SNP array (Supplementary Fig. 5). Of the four existing CNA methods used to analyze this tumor, BIC-seq produced a similar CNA profile, while the remaining three methods detected only a single CNA (Supplementary Fig. 5a,b). CONSERTING identified 12 SVs matching the boundaries of amplification CNAs (Supplementary Fig. 5c), all of which were validated experimentally by PCR followed by Sanger sequencing across the breakpoints[1] (Supplementary Fig. 5d). By contrast, standalone CREST analysis could only identify 5 SVs (Supplementary Fig. 5c). FISH was performed to validate two CNAs in this region, one identified by both CONSERTING and SNP array (blue probe on Supplementary Fig. 5c), and the other found exclusively by CONSERTING (red probe on Supplementary Fig. 5c). The FISH image showed co-amplification of the two regions in 50-70% of the tumor cells[1] (Supplementary Fig. 5e), which is consistent with the weak SV signature and the 0.5-fold amplification result reported by CONSERTING. SVs in this region satisfied criteria for chromothripsis[2] and formed a fully connected graph expected from a chromothripsis event (Supplementary Fig. 5c), which projects that the red and blue FISH probes are 2.2Mb apart if all 7 SVs in between occurred *in cis*. However, the fluorescence signals of these two probes are distinctively separated (Supplementary Fig. 5e), indicating that not all SVs are *in cis*[1].

Localized SV analysis implemented in CONSERTING is able to improve the sensitivity of both CNA and SV detection as demonstrated in the two tumors with subclonal CNAs and SVs that

match the pattern of chromothripsis (SJRB003 and SJLGG039). All SVs detected by CONSERTING in these two tumors were experimentally verified by orthogonal sequencing[1, 3]. While the novel CNAs were verified by 3-color FISH in both tumors, co-occurrence of multiple SVs on the same chromosome was confirmed in the LGG tumor but not in the RB tumor, demonstrating the challenges of determining the long-range phase of structural variations derived from WGS data. Multiple subclonal fusion signals detected in the LGG tumor by 3-color FISH (Fig. 3d) also show that a subset of tumor cells may have undergone additional re-arrangments after the complex event generating multiple *cis* SVs.

Supplementary Data 3

Novel CNAs in three unpublished leukemia WGS samples

The complete genomic landscapes of three ALL samples, SJTALL015 (T-ALL), PALETF and PALJDL (B-ALL) have not previously been reported. 10 novel CNAs were identified by CONSERTING but not by SNP array in gene coding regions, 9 of which were validated experimentally. The most notable was a 10kb deletion in *NOTCH1* predicted to result in loss of the negative regulatory region (NRR). Loss of NRR is expected to cause successive ligand-independent cleavages of the protein producing an active, intracellular *NOTCH1* (ICN). In this study we are able to confirm the expression of the mutant transcript as well as the presence of ICN in tumor. Previous studies have identified 5' and intragenic deletions of Notch1-driven mouse T-ALL[4] and 5' deletions in human T-ALL[5], but no intragenic deletions have previously been identifed in human T-ALLs by SNP array analysis. Our study indicates that the previous negative finding in the human disease is likely to be a consequence of lower resolution and poor coverage of *NOTCH1* by SNP arrays. The novel discovery of a 5' deletion of *NOTCH1* in human cases also reveals potential limitations of a proposed strategy of targeting *NOTCH1* with inhibitory antibodies[6] that recognize either the ligand-binding domain or NRR, as tumors lacking the NRR domain will be unable to respond to such therapy. Protease inhibitors and novel agents that target Notch in the nucleus may be a more prudent alternative for patients with such deletions.

To investigate whether there could be additional novel exonic CNAs missed by CONSERTING, we randomly selected 8 novel deletions affecting gene coding exons predicted by SegSeq but not CONSERTING from one of the three cases (PALJDL) for experimental validation (Supplementary Table 6). These 8 deletions spanned from 2.8-12.2 kb and none were validated.

Supplementary Data 4

Investigation of “fractured” genome in PCGP and TCGA WGS data

Of the 20 pediatric WGS data analyzed in this study, the initial WGS data generated for SJRB002 produced the highest number of CNA segments: a total of 16,920, 16,830, 1,826 and 14,398 CNAs were predicted by CONSERTING, BIC-seq, CNV-Seq and SegSeq, respectively. The massive number of CNAs predicted from WGS data makes the tumor genome appear “fractured” (Supplementary Fig. 8a,c). However, the CNA boundaries did not have matching SVs from structural variation analysis performed in CONSERTING. To further investigate the “fractured” genome, we made a new sequencing library and performed a second WGS at 10x coverage. Amplification of 6p is the only CNA that was replicated in the original and the new WGS data (Supplementary Fig. 8b,d), suggesting that the “fractured” genome derived from the first WGS data was artifactual. This pattern of “fractured” genome also exists in the TCGA GBM whole-genome sequencing data which was not replicated in the CNA profile generated by SNP-array (Supplementary Fig. 8e,f).

Supplementary References:

1. Zhang, J. et al. *Nature* **481**, 329-334 (2012).
2. Korbel, J.O. & Campbell, P.J. *Cell* **152**, 1226-1236 (2013).
3. Zhang, J. et al. *Nat Genet* **45**, 602-612 (2013).
4. Ashworth, T.D. et al. *Blood* **116**, 5455-5464 (2010).
5. Haydu, J.E. et al. *Blood* **119**, 5211-5214 (2012).
6. Wu, Y. et al. *Nature* **464**, 1052-1057 (2010).

Supplementary Tables

Supplementary Tables 2 – 6 are in the files Supplementary_Table_2.xls, Supplementary_Table_3.xls, Supplementary_Table_4.xls, Supplementary_Table_5.xls and Supplementary_Table_6.xls

Supplementary Table 1 Sample characterstics of 43 tumor pairs.

All except for COLO-829 were derived from patient tumors. Leukemia samples have >= 70% blast. The tumor purity for retinoblastoma is estimated to be 75-100% and SJLGG039 to be 30% based on the mutant allele fraction derived from deep sequencing. All except for COLO-829 were assayed by Affymetrix Genome-Wide Human SNP Array 6.

*Cases with num. of tumors =2 have tumors from diagnosis (D) and relapse (R). Coverage for these cases included germline (G), Diagnosis (D) and Relapse (R).

#COLO-829 sample used in this study was obtained by mixing the tumor (COLO-829) and normal (COLO-829BL) genomic DNA in equal concentrations. The estimated tumor purity based on SNV mutant allele fraction is 40%. The CNA of the un-diluted tumor cell line was assayed by SKY map (http://www.path.cam.ac.uk/~pawefish/OtherCellLineDescriptions/COLO829.html).

Cases	Source	*Num. Tumors	Tumor Type	Gender	Race	Age at diagnosis	Genome Coverage (G/D) or (G/D/R)
Pediatric							
SJTALL001	PCGP	1	Leukemia	M	B	10	26/38
SJTALL002	PCGP	1	Leukemia	F	W	6	24/34
SJTALL003	PCGP	1	Leukemia	M	B	7	25/32
SJTALL004	PCGP	1	Leukemia	M	W	3	26/32
SJTALL005	PCGP	1	Leukemia	F	B	12	27/32
SJTALL006	PCGP	1	Leukemia	M	W	10	26/30
SJTALL007	PCGP	1	Leukemia	M	W	18	28/36
SJTALL008	PCGP	1	Leukemia	M	W	2	28/32
SJTALL009	PCGP	1	Leukemia	M	B	12	26/32
SJTALL011	PCGP	1	Leukemia	F	W	8	24/34
SJTALL012	PCGP	1	Leukemia	F	W	15	28/35
SJTALL013	PCGP	1	Leukemia	F	B	13	29/34
SJTALL015	PCGP	1	Leukemia	M	W	7.4	45/38
SJRB001	PCGP	1	Retinoblastoma	M	W	2.3	25/29
SJRB002	PCGP	1	Retinoblastoma	F	W	0.8	27/28
SJRB003	PCGP	1	Retinoblastoma	M	B	1.1	28/35
SJRB004	PCGP	1	Retinoblastoma	F	W	1.8	27/33
SJLGG039	PCGP	1	Low-grade glioma	F	W	10	37/64
PALJDL	COG	1	Leukemia	M	Hispanic	3.2	27/29
PALETF	COG	1	Leukemia	F	W	7.6	28/27
Adult							
TCGA-14-1034	TCGA	2	Glioblastoma	F	NA	60	46/63/66
TCGA-06-0171	TCGA	2	Glioblastoma	M	W	65	47/72/72
TCGA-19-1389	TCGA	2	Glioblastoma	M	W	51	33/68/65
TCGA-06-0211	TCGA	2	Glioblastoma	M	W	47	49/67/68
TCGA-06-0190	TCGA	2	Glioblastoma	M	W	62	31/65/65
TCGA-06-0125	TCGA	2	Glioblastoma	F	W	63	45/67/69
TCGA-14-1402	TCGA	2	Glioblastoma	F	W	58	41/63/65
TCGA-06-0210	TCGA	2	Glioblastoma	F	W	72	40/64/64
TCGA-19-5960	TCGA	1	Glioblastoma	M	W	56	37/37
TCGA-27-1831	TCGA	1	Glioblastoma	M	W	66	35/38
TCGA-06-0152	TCGA	1	Glioblastoma	M	W	68	36/39
TCGA-06-0145	TCGA	1	Glioblastoma	F	W	53	53/52
TCGA-06-0185	TCGA	1	Glioblastoma	M	W	54	31/32
TCGA-06-0648	TCGA	1	Glioblastoma	M	W	77	60/38
#COLO-829	PCGP	1	Melanoma	M	W	45	53/53

Supplementary Table 2

Manually curated autosomal CNAs by SNP array analysis in 12 ETP-TALL tumors.

Supplementary Table 3

Comparison of CNAs from whole-genome sequencing by CONSERTING, BIC-seq, SegSeq, CNV-Seq, FREEC with manually curated CNVs from SNP array analysis in the 12 ETP-TALL samples.

Supplementary Table 4

Comparison of CNAs from whole-genome sequencing by CONSERTING and BIC-seq with CNAs from SNP array analysis in the 22 TCGA_GBM samples.

Supplementary Table 5

CNA profile and SV matching status of diluted COLO-829 sample analyzed by CONSERTING

Supplementary Table 6

Experimental validation of novel CNAs in coding exons identified by CONSERTING and SegSeq in the two COG samples and SJTALL015.