Supplementary information

Table of Contents

Table of Supplementary Figures and Tables

Supplementary results

1. The Abnormality

a. Background

Intrachromosomal amplification of chromosome 21 (iAMP21) was originally identified as a distinct cytogenetic subgroup of childhood acute lymphoblastic leukemia (ALL) in 2003[1,2], following reports of a number of sporadic cases[3-12]. In all studies, iAMP21 patients had B-cell precursor ALL (BCP-ALL), were older (median age of 9-11 years), and generally had low white cell counts (WCC). Prospective screening in recent childhood trials has determined the incidence to be 2%[13]. A significant finding was that patients with iAMP21 had an inferior outcome when treated on standard therapy, compared to other patients treated on the same protocols[14-16]. However, treatment as high-risk on intensive therapy was shown to dramatically decrease relapse risk and improve survival[13,17]. iAMP21 was defined as a primary cytogenetic abnormality with a complex structure of one copy of chromosome 21, comprising multiple regions of gain, amplification, inversion and deletion, identified from cytogenetics, fluorescence *in situ* hybridization (FISH) and genomic analysis, which was highly variable between patients[18-20]. We identified a common region of highest level amplification spanning 5.1Mb of chromosome 21 from 32.8-37.9Mb, within which the *RUNX1* gene is located[18]. We proposed that the abnormal chromosome 21 arose through a breakage-fusion-bridge (BFB) mechanism[19], supported by the observation of anaphase bridges involving chromosome 21 in some iAMP21 patients[21]. Further studies pointed to clustered breakpoints within the *PDE9A* gene in some patients, with complex events around microhomology-mediated end joining as preceding or initiating the BFB cycles[22]. FISH using probes directed to *RUNX1* provides the most reliable detection method[23]. Three or more extra copies of *RUNX1* on a single abnormal chromosome 21 (a total of 5 or more *RUNX1* signals per cell) define iAMP21, a definition that has now been adopted internationally[24].

b. Risk of rob(15;21)(q10;q10)c developing iAMP21 ALL

Robertsonian translocations are whole arm translocations between the short arms of the acrocentric chromosomes (13-15, 21 and 22), found in ~1 in 1,000 newborns[25,26]. Of these, rob(15;21)c is exceedingly rare, accounting for only 0.5-1% of all Robertsonian translocations.

The annual incidence of childhood (0-14 years) ALL is 35 cases per million in England and Wales[27]. The frequency of iAMP21 in childhood ALL is 2.1%[28], thus the annual incidence of iAMP21 ALL in children (0-14 years) is 0.74 cases per million. We have shown for the first time that the incidence of rob(15;21)c among iAMP21 is 3.2%, making the frequency of rob(15;21)c in childhood ALL 0.07% at an incidence of 0.02 per million with a ~2700 increased risk of developing iAMP21 ALL (Supplementary Table 1). Interestingly all cases of rob(15;21)c with ALL were iAMP21, highlighting the predisposing nature of rob(15;21)c to the development of iAMP21.

To endorse this specificity, within the same ALL cohort we found only two other constitutional Robertsonian translocations involving chromosome 21:

rob(14;21)(q10;q10)c, with no evidence of iAMP21 in their acquired karyotype. In addition, we interrogated the cytogenetics database from the Munich Leukemia Laboratory. Among 67,000 referrals for hematological malignancies, only two patients were identified with rob(15;21)c, one was a lymphoma in which the acquired karyotype was not available and one was a reactive node with no evidence of cancer. Among 26,000 referrals for hematological malignancies to the West Midlands Regional Genetics Laboratory, Birmingham Women's Hospital, UK, a single patient with rob(15;21)c was found over a 25 year period and this patient developed der(15;21) and was included in this study. Sixteen cases of rob(15;21)(q10;q10)c emerged from approximately 300,000 referrals to UK regional cytogenetics laboratories in Birmingham, Manchester, Newcastle, Salisbury and Sheffield for investigations of infertility or birth of a previous Down syndrome child but no evidence of malignancy. These observations confirmed the extreme rarity of rob(15;21)(q10;q10)c and indicate that although carriers of rob(15;21)c have a strong predisposition to iAMP21 ALL, a number of carriers also exist who will not develop iAMP21 ALL.

Supplementary Table 1: Relative risk of iAMP21 ALL in carriers of constitutional Robertsonian translocation: rob(15;21)c.

	rob(15;21)c	**General population of children 0-14 years of age**
iAMP21 ALL	rob(15;21)c iAMP21 ALL = 3.2% of iAMP21 ALL $\approx 0.02/10^6$	Childhood ALL $= 35/10^6$ iAMP21 ALL = 2.1% of childhood ALL $\approx 0.74/10^6$
No ALL	rob(15;21)c $\approx 10/10^6$ births	≈1

Relative risk = $(0.02/10) / (0.74 \times 10^{-6})$

≈ 2700

2. The Cohort

The cohort includes sporadic iAMP21 patients (n=21) and iAMP21 with rob(15;21)(q10;q10)c (n=12) (Supplementary Table 2). The sporadic iAMP21 cases were those with DNA available for sequencing/SNP arrays among the 108 UK patients reported in the Ponti di Legno International Childhood ALL Group study, which includes a total of 530 iAMP21 patients[29]. Four of the 12 rob(15;21) cases in this study are also reported in this International study. The remainder were collected through International collaboration. The patient numbers are indicated in Supplementary Table 2 for cross-reference.

The median age of the sporadic iAMP21 patients was 9 years, as also shown in our previous studies[13]. The median age of the der(15;21) iAMP21 patients was younger at 6 years (range 3-13 years), although the number of patients is small. The karyotypes of this cohort are indicated. Interestingly two of the der(15;21) iAMP21 cases showed near tetraploid karyotypes (patients 25124 and 25190). All of the sporadic iAMP21 cases were defined by FISH and SNP6.0 data were available (Supplementary Figure 1A) whereas SNP6.0 data were available for only 5 of the der(15;21) iAMP21 patients (Supplementary Figure 1B). While the historical cases had no FISH data, they were included in the table as the karyotypes of the leukemic cells indicated similar abnormalities of chromosomes 15 and 21 as the other der(15;21) iAMP21 cases. The karyotypes of patients 86 and 25124 were only seen at relapse, as their diagnostic karyotypes failed.

The sequencing studies were approved by the Cambridgeshire Local Research Ethics Committee. Written informed consent was obtained from the patients' guardians.

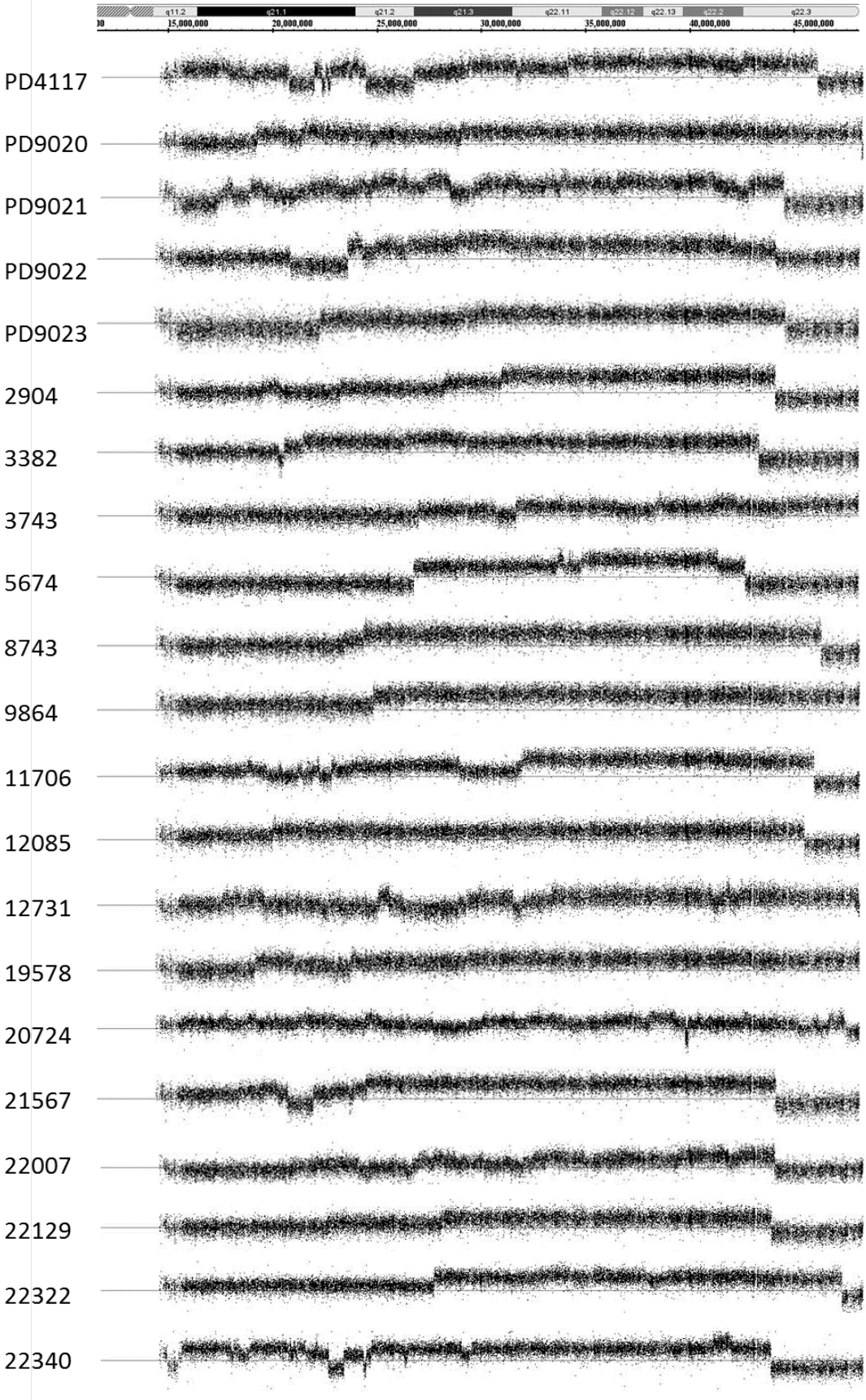
Supplementary Table 2: Clinical, demographic and cytogenetic data of iAMP21 patients included in the study.

ID	Age (yr)	Sex	WBC ($x10^9$)	Abnormal Karyotypes	Morphology of iAMP21
Patients with sporadic iAMP21 sequenced					
PD9023a	5	M	31	47,XY,+X,del(16)(q13),i(17)(q10),ider(21)(q10)dup(21)(q?)[3]/ 47,idem,add(7)(p1?)[3]	iso
PD9022a	9	F	1.7	46,XX[45]	UK
PD4117a	10	F	NA	47,XX,+10,der(21)r(21)(q?)dup(21)(q?)[10]	ring
PD9020a	7	F	9.1	47,XX,+X,-21,+mar[15]	UK
PD9021a	11	F	7.1	46,X,der(X)add(X)(p2?)add(X)(q2?),der(21)r(21)(p1?q2?)dup(21)(q22)[9]	ring
Patients with der(15;21) sequenced					
PD7171a	3	M	84.2	43~44,XY,del(5)(q11q13),der(15;21)(q10;q10)dup(15;21)(q?;q?),del(16)(q22), -20[cp5]/45,XY,rob(15;21)(q10;q10)c	UK
PD10008a	5	F	10.4	48,XX,add(1)(q?),del(13)(q1?),der(15;21)(q10;q10)dup(15;21)(q?;q?)x2,+22,+ mar[2]/45,XX,rob(15;21)(q10;q10)c	iAMP21x2
PD7170a	10	M	8.3	46,XY,der(1)t(1;13)(q2?5;q12),del(9)(p2?1),-11,-13,-15,del(16)(q10),-17,- 21,+3mar[5]/45,XY,rob(15;21)(q10;q10)c	UK
PD10009a	6	M	NA	45,XY,+X,der(15;21)r(15;21)(?;?),-20[9]/46,idem,+21[4]/ 45,XY,rob(15;21)(q10;q10)c	ring
Other der(15;21) patients					
86	4	M	4	44,XY,-7,der(15;21)dup(15;21)(q?;q?)[10]/45,XY,rob(15;21)(q10;q10)c	iso
24453	4	M	NA	47,XY,+X,?i(3)(p10),del(6)(q22q24),del(7)(q31q35),+10,+18,+21,der(21)i(q10)i ns(21;?)(q21;?),add(22)(p13)[cp9]/45,XY,rob(15;21)(q10;q10)c	UK
25010	6	M	NA	49,XY,+der(X)t(X;15)(p11.2;q22),der(15;21)(q10;q10)del(15)(q22)x2,+mar[6]/ 45,XY,rob(15;21)(q10;q10)c	iAMP21x2
25124	NA	M	NA	84-85<4n>,XXYY,add(1)(p36)x2,-2,-2,-4,-,4,-7,-7,+10,-12,-12,-15,-15,+16,+19,- 21,-21,+2mar[cp4]/45,XY,rob(15;21)(q10;q10)c	iAMP21x2
25155	13	M	NA	46,XY,inv(3)(?q?q),+13,-15,-21,+r[3]/45,XY,rob(15;21)(q10;q10)c	ring
25190	8	F	NA	91<4n>,XXXX,-7,-8,-9,+10,der(15;21)(q10;q10)x2,-20,ider(21)(q10)x6[8]/ 45,XX,rob(15;21)(q10;q10)c	iAMP21x2
25631	11	M	NA	46,XY,t(12;14;19)(p13;q13;q13.3),del(13)(q12q21),- 18,i(21)(q10)x2/45,XY,rob(15;21)(q10;q10)c	iAMP21x2
85631	10	F	NA	45,XX,der(?)der(15;21)(q10;q10),?del(19)(p13p13),?t(?;19)(?;p13)[5]/ 45,XX,rob(15;21)(q10;q10)c	
Other sporadic iAMP21 patients					
2904	15	F	1.3	46,XX,del(7)(q22q32),dup(21)(q?)[16]	UK
3382	11	M	6.1	46,XY,i(9)(q10),del(11)(q2?1),dup(21)(q?)	UK
3743	7	F	15.2	45,XX,dup(8)(p?),-11,der(15)t(11;15)(?;q24),der(21)r(21)(q?)dup(21)(q?)[2]	ring
5674	6	M	55.1	47,XY,+X,dup(21)(q?)[7]	UK
8743	8	F	25.9	46,XX,t(7;9;17)(q22;p1?;p1?),del(11)(q23q2?5),dup(21)(q?)[7]	UK
9864	10	M	12.7	47,XY,dup(21)(q?),+dup(21)(q?)[6]	iAMP21x2
11706	5	M	5.8	Fail	UK
12085	9	M	1.1	46,XY,del(13)(q14),del(16)(q12.1),dup(21)(q?)[9]	UK
12731	13	M	2.3	47,XY,+21[18]	UK
19578	11	M	1.5	48,XY,+X,t(6;20)(p1?;q1?),t(7;9)(p1?;p2?),i(9)(q10),+12,dup(21)(q?)[9]	UK
20724	3	M	5.4	Fail	UK
21567	8	F	4.9	52,XX,+9,-12,-21,+7mar[cp4]	UK

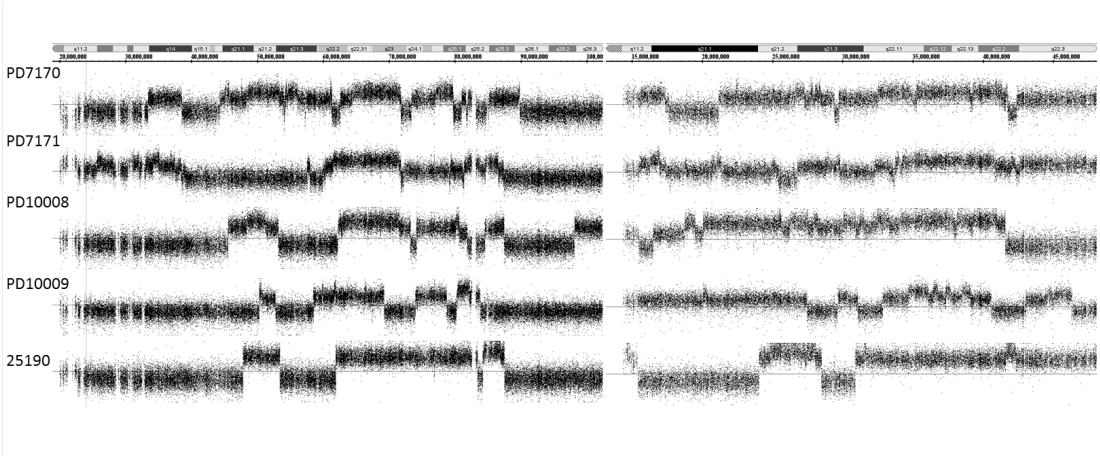
22007	11	M	4.3	46,XY,add(21)(p11)[4]	UK
22129	8	F	15.4	46,XX,t(1;5)(p3?2;q3?1),add(21)(q22)[10]	UK
22322	8	F	10	46~47,XX,+X,add(7)(p22),del(11)(q23),del(13)(q12q14),dup(21)(q22),+r[cp8]	ring
22340	10	F	10.6	46,XX,ider(21)(q10)inv dup(21)(q1?q2?)[9]	UK

Iso- isochromosome, iAMP21x2- duplication of iAMP21 chromosome, NA- not available, UK-unknown.

A


q11.2
q21.1
q21.2
q21.3
q22.11
q22.12
q22.13
q22.2
q22.3
15,000,000
20,000,000
25,000,000
30,000,000
35,000,000
40,000,000
45,000,000
PD4117
PD9020
PD9021
PD9022
PD9023
2904
3382
3743
5674
8743
9864
11706
12085
12731
19578
20724
21567
22007
22129
22322
22340

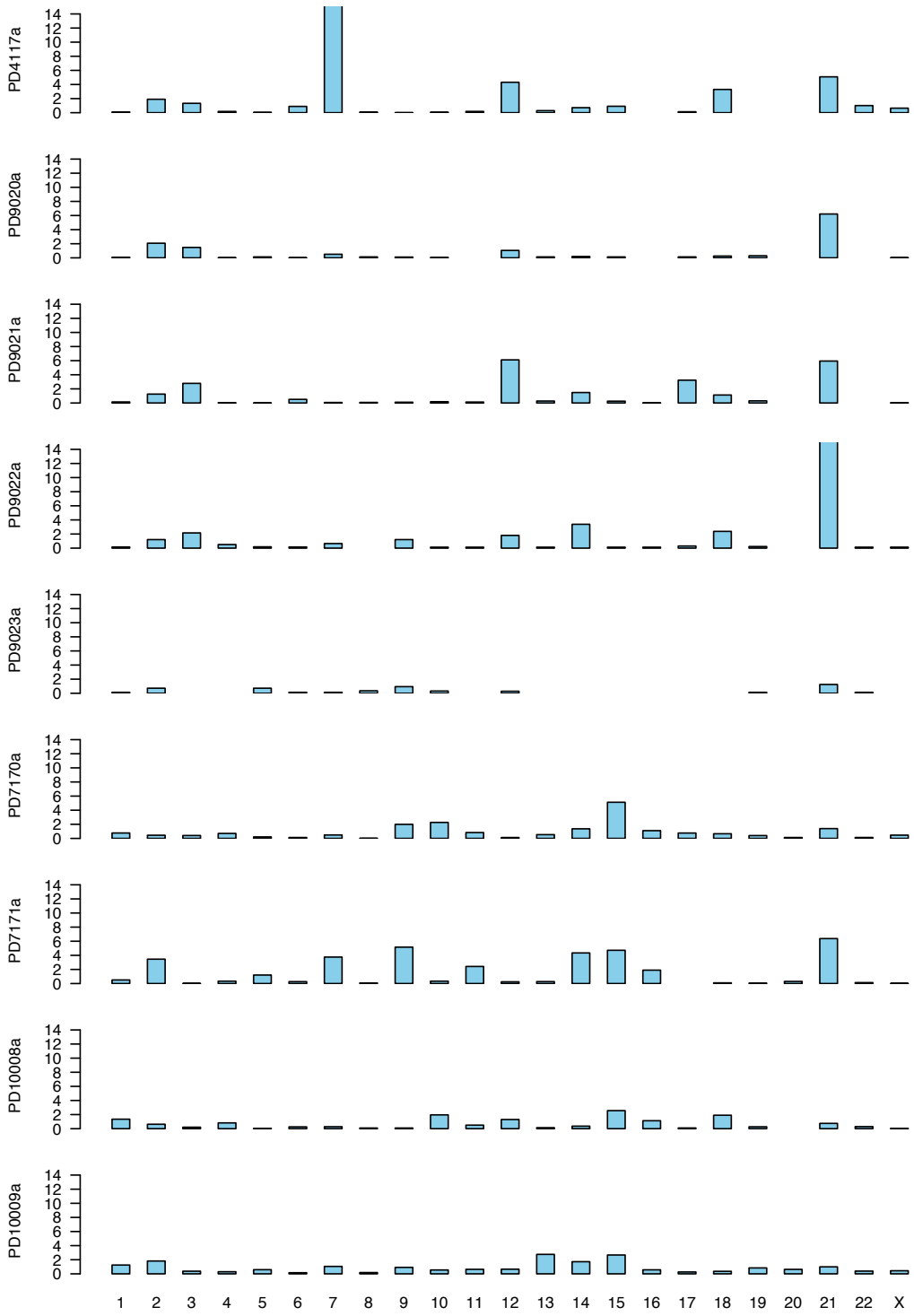
B



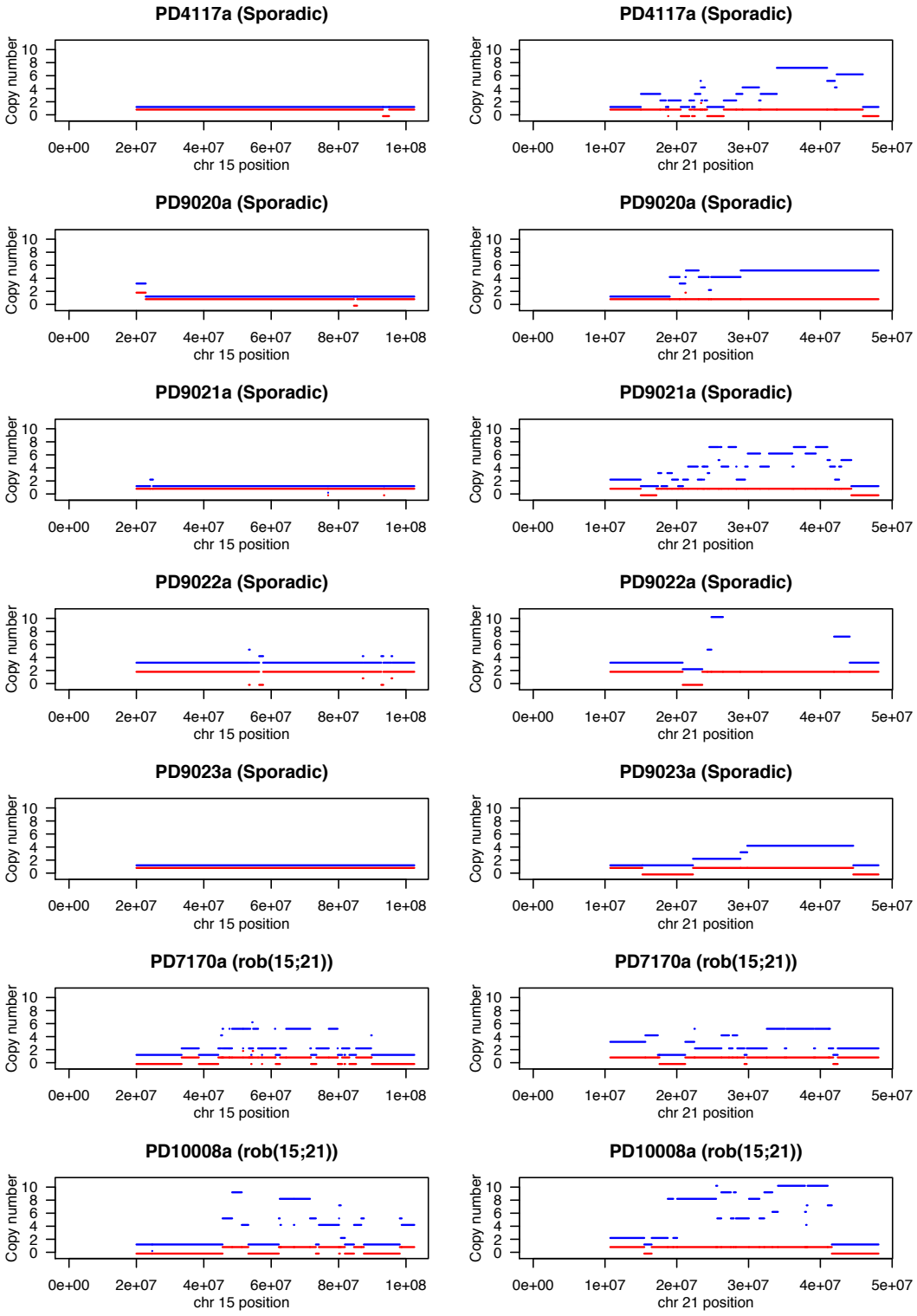
Supplementary Figure 1: SNP6.0 array profiles of chromosome 21. (A) SNP6.0 profiles from the 21 sporadic iAMP21 cases. (B) SNP6.0 array profiles of chromosomes 15 (left) and 21 (right) from 5 der(15;21) iAMP21 ALL.

3. Inferring chromothripsis

Recently Korbel and Campbell have proposed six genomic hallmarks of chromothripsis[30]. To evaluate the evidence for chromothripsis in our samples, we assessed the derivative chromosomes 21 and der(15;21) against these criteria. Five of the criteria were applicable to our samples. Out of these, all samples were consistent with the criterion of random rearrangement orientation, and varying numbers of samples conformed to the other applicable criteria (Extended Data Table 1).



Supplementary Figure 2: Several iAMP21 chromosomes show evidence for breakpoint clustering. Statistical significance for deviation from the null hypothesis of no rearrangement breakpoint clustering for each sample and chromosome. Chromosomes are shown on the X-axis and Y-axis shows $-\log_{10}$(p-value). At *p*-value threshold 0.001, 5/9 chromosomes 21 and 2/4 chromosomes 15 (Robertsonian cases) show evidence of non-random rearrangement breakpoint distribution.



Supplementary Figure 3: Rearrangements on chromosome 21 are limited to the iAMP21 allele. Allele specific copy number analysis on samples for which matched normal SNP6.0 data was available, showing ubiquitous minor allele copy number of 1. Blue and red segments show major and minor copy numbers, respectively.

4. The Mechanism

A. Introduction

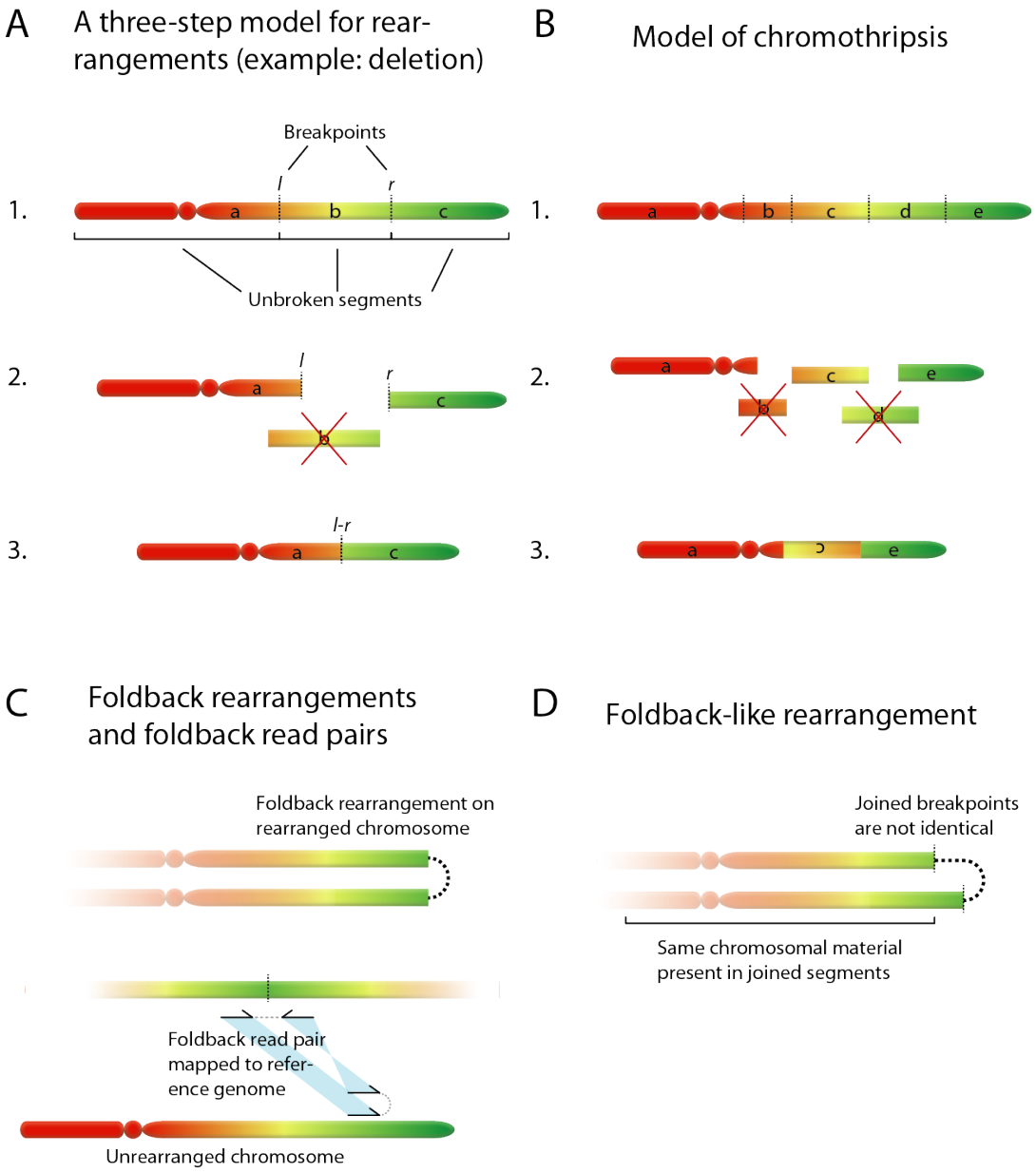
Massively parallel sequencing data provide a rich source of information about the chromosomal architecture of cancers, as both copy number and genomic rearrangements can be extracted at high resolution. However, much of this information is complex and non-trivial to interpret because cancer-associated chromosomes are disturbed over megabase scales, beyond the linkage that can be achieved by short sequencing reads. Here we describe a framework for inferring the temporal evolution of the chromosome specific rearrangements of iAMP21 ALL from massively parallel sequencing data. The framework is built on core assumptions, from which structural rules of specific rearrangements can be derived. Such rules can then be used to deduce the nature and ordering of historical rearrangements that have occurred on sequenced chromosomes.

In section B, we describe our model for several types of complex rearrangement processes. We then define rearrangement metrics that are informative for inferring temporal evolution. In section C, we describe the base assumptions that underlie our deductions. In section D we derive a set of tenets that link specific rearrangements to the corresponding patterns in rearrangement metrics. In section E, we challenge the tenets of section D and show that they are consistent with simulated data. In section F, we apply the tenets of section D to deduce the rearrangement history of the iAMP21 chromosome in the samples sequenced in this study.

B. Definitions

A rearrangement is modeled as a three-step process, where: (1) a chromosome is broken at any number of breakpoints, generating discrete chromosomal fragments; (2) individual fragments are removed, retained or duplicated once; and (3) the retained and duplicated fragments are joined back together. For example, a deletion is modeled as generation of two breakpoints, *l* and *r*, on a chromosome, followed by loss of the fragment demarcated by *l* and *r* and joining of the two breakpoints into a derivative chromosome (Supplementary Figure 4A).

In chromothripsis, several breakpoints are first generated, producing individual discrete chromosomal fragments. Some of the fragments are lost, and the remaining fragments are joined together in random order and orientation into a contiguous derivative chromosome (Supplementary Figure 4B)[30]. In BFB cycles, two duplicated copies of a chromosomal region are joined together through a fold-back inverted rearrangement at the shared double strand DNA break (Supplementary Figure 4C). Read pairs that span such rearrangements are called fold-back read pairs (Supplementary Figure 4C). A fold-back-like rearrangement is similar to fold-back rearrangement, but the two inverted segments do not share an exactly identical breakpoint (Supplementary Figure 4D).



Supplementary Figure 4: Basic assumptions of the rearrangement model. A: A three-step model used to model rearrangements. First, the affected chromosome is broken through breakpoints into fragments (1). Fragments are then possibly deleted or duplicated once (2). Finally, any retained fragments are joined together through their breakpoints in the orientation dictated by the modeled rearrangement event (3). Under the assumption of unique breakpoints, no further breakpoints can be generated exactly at breakpoint *l* or *r* in any later rearrangement events. B: A generative model for chromothripsis. A chromosome breaks into (a large number of) fragments (1), after which each fragment is randomly lost or retained (2), with retained fragments joined together in random order and orientation. C: Fold-back rearrangements are defined as rearrangements that join breakpoints of directly inverted duplications. D: Fold-back-like rearrangements are head-to-head or tail-to-tail inverted rearrangements with no intervening copy number changes or

breakpoints. The two inverted copies are not joined from an identical duplicated breakpoint.

Below we define the metrics used for inference of rearrangement history. The process of gathering and representing these metrics is illustrated in Extended Data Figure 3.

Copy number state distribution	The frequency distribution of copy number states over the chromosome.
Copy number step	The magnitude of change in copy number at each breakpoint location as the chromosome is traversed in its reference configuration (that is, first to last base-pair of chromosome 21). The size of a copy number step is defined as its absolute value.
Copy number jump	The magnitude of change in copy number between the two segments joined by a rearrangement, defined as the copy number at the low end of the rearrangement join subtracted from the copy number at the high end of the rearrangement join. For example, if locus *A* in chromosome 1 were joined to locus *B* in chromosome 2 through a rearrangement, the associated jump size would be |copy number at *B* – copy number at *A*|. The size of a copy number jump is defined as its absolute value.
Copy number trajectory	A graph that shows the collapsed copy number segments and their levels across a chromosome. In this graph, every copy number segment is represented as one point regardless of the size of the segment, in order to emphasize the changes in copy number between adjacent segments.

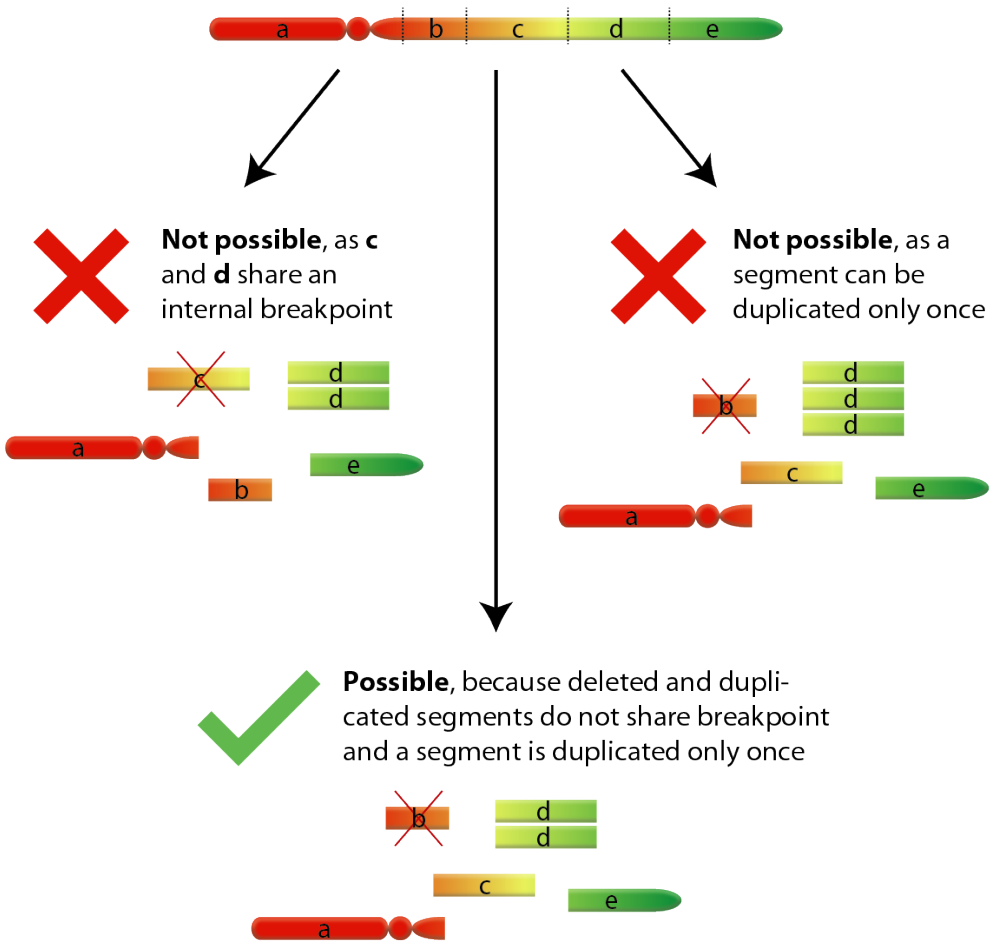
C. Assumptions

Throughout the study, we apply the **assumption of unique breakpoints** (also known as the 'infinite sites assumption'), which states that the probability of two independent internal breakpoints occurring at the exact same chromosomal position is zero. For example, if a segment between breakpoints *l* and *r* is deleted generating the rearrangement join *l*-*r*, the derivative chromosome is assumed to not break at either point *l* or *r* again in any subsequent rearrangements (Supplementary Figure 4A). During a single rearrangement, a deleted and a duplicated segment never share the same breakpoint (left hand side in Supplementary Figure 5).

Note that this assumption applies only to internal breakpoints. A chromosome can first undergo whole-chromosome duplication followed by whole-arm deletion, in

which case a terminal (or telomeric) breakpoint is shared by the duplication and deletion events.

We also apply the **assumption of single duplication**, whereby any segment may duplicate only once in a rearrangement event (right hand side in Supplementary Figure 5). Of course, a chromosomal region may be amplified multiple times, but the assumption dictates that each round of amplification must occur through the duplication of different (but overlapping) segments, none of which shares an identical breakpoint with another segment. We note that this assumption is erroneous for circular double minute chromosomes that may be duplicated through rolling circle replication. However, since cytogenetic analysis established that none of the samples in this study had double minutes, we retain this assumption (Supplementary Figure 5).



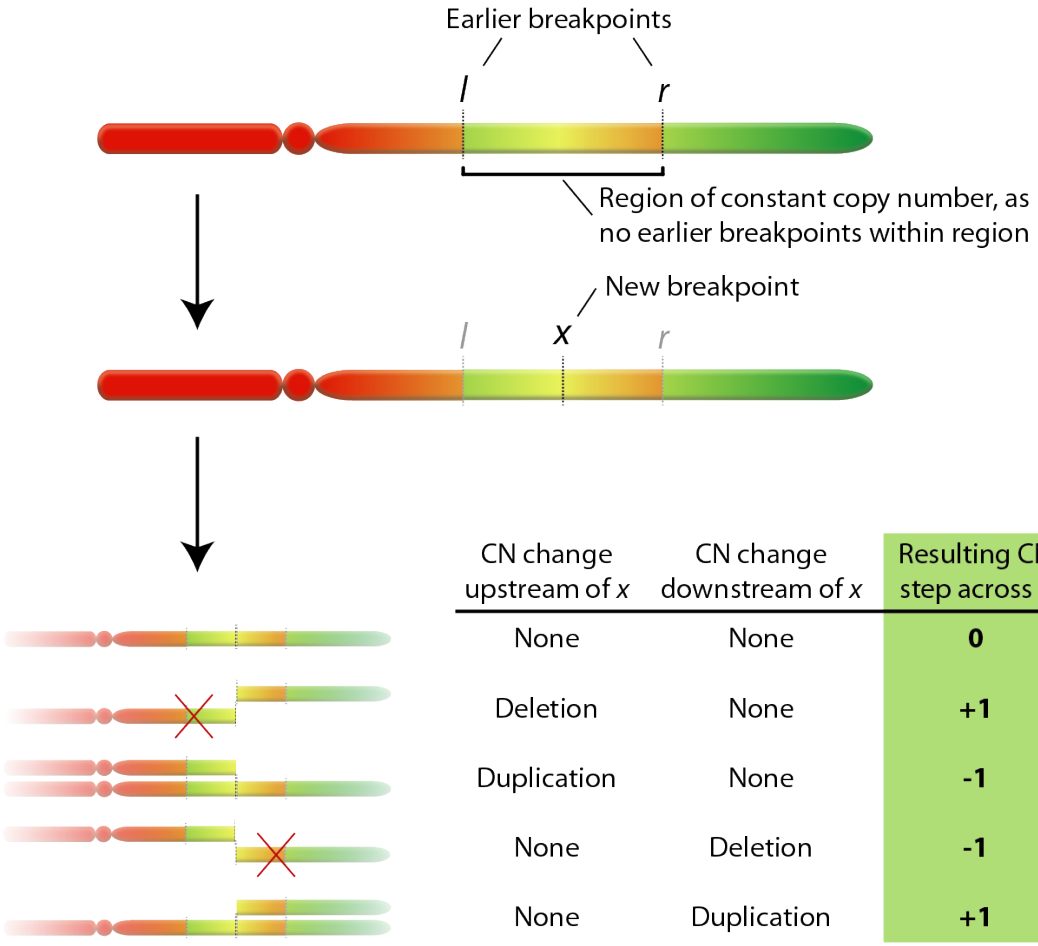
Supplementary Figure 5: Illustration of the assumption of unique breakpoints and the assumption of single duplication. A deleted segment cannot be directly adjacent to a duplicated segment (left arrow). A segment can duplicate at most once during a rearrangement (right arrow). Otherwise segments can be deleted or duplicated freely (middle arrow).

D. Inferring the nature and order of sequences of rearrangements

Tenet 1. Rearrangements generate copy number steps of 0, -1 or 1

Let us study all the possible copy number steps generated over a new breakpoint at position *x*. Because of the **assumption of unique breakpoints**, position *x* is distinct from any previously generated breakpoints. We label the previously generated breakpoint closest to *x* from the p-telomeric side as *l*. If there are no such breakpoints, then we label the p-telomere as *l*. Similarly, we label the closest previous breakpoint to *x* from the q-telomeric side as *r*. Since there are no previous rearrangement breakpoints between *l* and *r*, the region between *l* and *r* must have a constant copy number (Supplementary Figure 6).

The breakpoint *x* divides the segment between *l* and *r* into two segments. Since we also assume that a deletion and a duplication event never share a breakpoint directly (**assumption of unique breakpoints**), the possible combinations of CN steps across *x* will be limited to 0, -1 and +1 (Supplementary Figure 6).

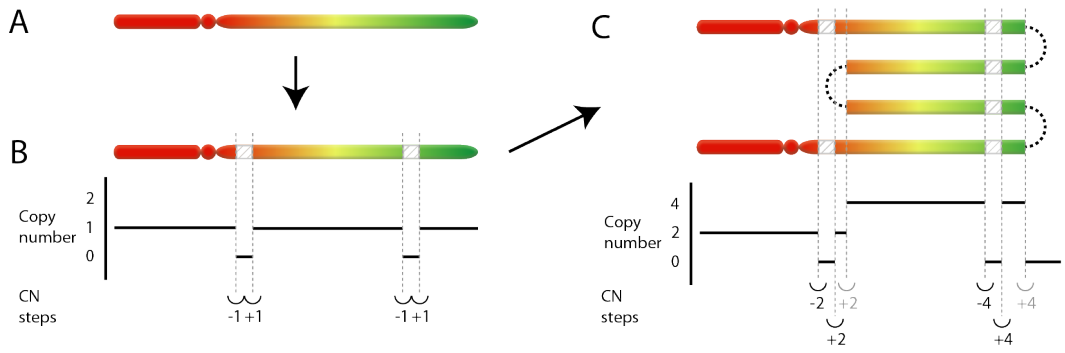


CN change upstream of *x*	CN change downstream of *x*	Resulting CN step across *x*
None	None	**0**
Deletion	None	**+1**
Duplication	None	**-1**
None	Deletion	**-1**
None	Duplication	**+1**

Supplementary Figure 6: New rearrangements can only generate copy number (CN) steps of 0, -1 or 1.

Tenet 2. CN steps of size >1 are caused by deletion and duplication events that overlap with previously generated rearrangement breakpoints.

Rearrangements separate regions that are adjacent in the reference genome. A subsequent duplication or deletion event may thus affect only one of two segments that are immediate neighbors in the reference genome. As a consequence, the CN change alters the CN of only one of the two adjacent segments, thereby altering the copy number step across the CN breakpoint (Supplementary Figure 7).



Supplementary Figure 7: An example of generation of CN steps of size >1. An unrearranged chromosome (A) undergoes two deletions, generating four CN segment breakpoints each with a CN step size 1 (B). Due to these rearrangements, subsequent CN altering rearrangement events (for example BFB events in this case) affect only the retained but not the deleted regions, exaggerating the size of the previously generated CN steps (C). In C, CN steps associated with original deletions are shown in black and CN steps associated with two BFB cycles are shown in grey.

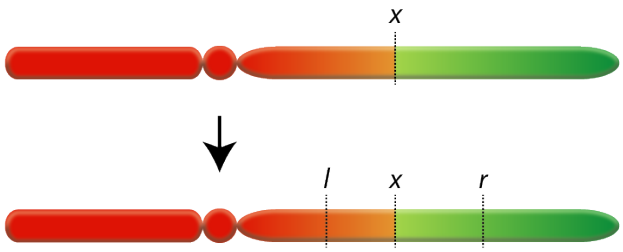
Tenet 3. Chromothripsis on an unrearranged, unduplicated region generates CN jumps of magnitude 0.

Since the affected chromosome is unrearranged and chromothripsis is assumed to be a non-duplicative process (Supplementary Figure 4B), all retained segments will have copy number 1. In the last stage of chromothripsis (Supplementary Figure 4B), the retained segments will be joined to each other generating (1 copy to 1 copy) CN jumps for every rearrangement join. Therefore all the generated CN jumps will have magnitude abs(1 - 1) = 0.

Please note that although chromothripsis (catastrophic chromosomal shattering followed by repair) itself is non-duplicative, a rearrangement process consisting of chromothripsis can as a whole be duplicative, if the rearrangement process involves duplicating steps, such as whole-chromosome duplication.

Tenet 4. The magnitude of a CN jump that starts as 0 will always remain 0 regardless of any subsequent rearrangements on the same chromosome.

Following the **assumption of unique breakpoints**, no further breakpoints will occur at the rearrangement joining point *x* from a previous rearrangement. Since the rearrangement at *x* joins two regions of copy number 1 together, no other copies of these regions are present elsewhere on the same chromosome. As a consequence, a subsequent CN change immediately upstream of *x* will always similarly affect the region immediately downstream of *x*, and vice versa. Thus the CN jump size across this rearrangement link will remain 0 (Supplementary Figure 8).



Earlier rearrangement break-point with CN jump from *k* copies to *k* copies across *x*

Later rearrangement with breakpoints *l* and *r* closest to *x* on each side

CN change at segment *l*-*r*	CN change at segment *l*-*x*	CN change at segment *x*-*r*	Resulting CN jump across *x*	Resulting CN jump size
None	None	None	*k* to *k*	*k* - *k* = **0**
Deletion	Deletion	Deletion	*k*-1 to *k*-1	(*k*-1) - (*k*-1) = **0**
Duplication	Duplication	Duplication	*k*+1 to *k*+1	(*k*+1) - (*k*+1) = **0**

Supplementary Figure 8: A CN jump from one copy to one copy will always remain as CN jump size 0 regardless of subsequent rearrangements on the same chromosome. Due to assumption of unique breakpoints, an earlier breakpoint *x* will always be within segments generated in subsequent rearrangements. Thus the copy number immediately upstream and downstream to *x* will always be altered to the same extent regardless of the nature of any subsequent rearrangements on the same chromosome.

Tenet 5. Rearrangement joins associated with chromothripsis on an unrearranged region have CN jumps of size 0 regardless of any subsequent rearrangements on the same chromosome.

Chromothripsis generates CN jumps from 1 copy to 1 copy on unrearranged chromosomal regions (Tenet 3). Therefore all the size of all associated CN jumps will remain 0 regardless of any subsequent rearrangements (Tenet 4).

Tenet 6. If chromothripsis occurred as an initiating event on an unrearranged chromosome, all the associated CN jumps will be of size 0 regardless of any subsequent rearrangements on the same chromosome.

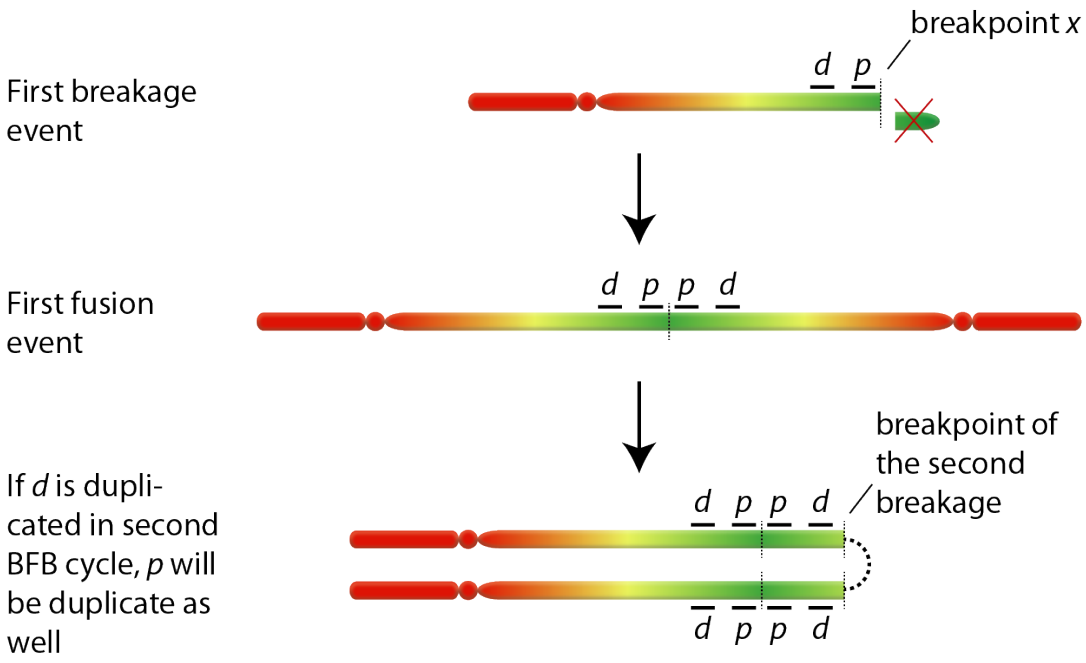
This follows directly from **Tenet 5** where the entire chromosome is now the unrearranged region.

(Weak) Tenet 7. Amplifying BFB cycles on a previously unrearranged chromosome generate a step-wise increasing CN pattern, which ends in a sharp decrease in CN demarcated by a fold-back rearrangement.

An amplifying BFB cycle is defined as a BFB cycle that results in amplification (as opposed to loss) of chromosomal material. This tenet is 'weak' because it does not apply strictly in all cases, although a statistical tendency towards the statement of the rule is present.

After the first BFB cycle, the initial telomeric breakpoint *x* that initiated the BFB cycle is at the center of the derivative chromosome. If any position *d* (for *distal* to *x*) is to be duplicated in the second BFB cycle, *x* and the surrounding region *p* (for *proximal* to *x*) will be duplicated as well due to this very structure (Supplementary Figure 9). Therefore at this stage copy number of *p* is larger or equal to copy number of *d*. Since the region near *x* will have the highest copy number in this case, *x* will also demarcate a steep CN decrease to the telomeric region lost during the first breakage event (Supplementary Figure 7). Furthermore, *x* is a fold-back rearrangement. Hence after only one or two BFB cycles, this tenet is strictly true.

With further BFB cycles, it is sometimes possible that the resulting chromosome has both positive and negative copy number steps when 'walking' the derivative chromosome in the reference configuration. However, since after two amplifying BFB cycles the proximal region around *x* has the highest CN, this region will be more likely to be duplicated in further rounds of BFB events as well. Hence with more than two BFB cycles, this tenet is not strictly true but rather a statistical tendency.



Supplementary Figure 9: Structure of a derivative chromosome following two BFB cycles. At the first BFB cycle, the chromosome is broken and then fused at breakpoint *x* while the region downstream of *x* is deleted. Two copies of *p* and *d* will be present in the fusion intermediate. During the second BFB cycle, if *d* is to be amplified, the chromosome has to be broken after the second copy of *d*. However, in this case both copies of *p* will be amplified as well. Thus if *d* is duplicated during the second BFB cycle, a region more proximal to the original breakpoint *x*, *p*, will be duplicated as well. As a result, after two amplifying BFB cycles, the CN of proximal regions to *x* will be at least as high as the CN of distal regions to *x*.

E. Applying the tenets of section D using simulated data

Scenario I: Relative ordering of BFB cycles and chromothripsis

Here we simulate two rearrangement events, two rounds of BFB and chromothripsis in their two possible orders. In both scenarios, we compare the rearrangement metrics introduced in section B to those predicted by the rules of section D. The predictions are contrasted between the two scenarios in Supplementary Table 3 and the simulated rearrangement metrics are shown in Extended Data Figure 4.

Supplementary Table 3: Predictions based on tenets in section D under the two different orderings of two rearrangement processes, 2 BFB cycles and chromothripsis.

Two BFB cycles then chromothripsis	Chromothripsis then two BFB cycles
Tenet 1 predicts that CN steps associated with chromothripsis are 0, 1 or -1.	Tenet 2 predicts that since some of the chromothripsis breakpoints are duplicated through BFB amplification, some chromothripsis associated CN steps will have size >1.
Tenet 6 does not apply, so chromothripsis associated rearrangement joins can have CN jumps ≠ 0.	Tenet 6 predicts that all chromothripsis associated rearrangement joins should have CN jump size of 0.
Tenet 7 predicts that CN pattern of the chromosome increases stepwise until a sharp drop in copy number demarcated by a fold-back rearrangement.	Tenet 7 does not apply so stepwise increasing CN pattern is not necessary.

Two BFB cycles generate only two breakpoints, and the majority of the breakpoints are generated by the chromothripsis event. In the scenario where the BFB cycles were the initiating events, all CN steps associated with chromothripsis were of size 1, and the CN jumps were often of size >0. In contrast, in the alternative scenario, most of the CN steps were of size >1, but none of the CN jumps involved differing copy numbers. Finally, in samples that had BFB cycles as the initiating event, a clear pattern of stepwise increase then fall in CN pattern was seen in their CN trajectories, but this pattern was absent in the alternative scenario. In conclusion, the observed rearrangement metrics from the two scenarios are consistent with the predictions based on the tenets in section D and listed in Supplementary Table 3 (Extended Data Figure 4).

Scenario II: Chromothripsis across three copy number states

In the der(15;21) iAMP21 ALL in this study, we saw an obvious rearrangement and CN pattern of chromothripsis. However, unlike the pattern of "conventional chromothripsis" (Supplementary Figure 4B), these chromosomes had a CN pattern oscillating among three copy number states. We therefore simulated two alternative scenarios that could generate such a CN and rearrangement pattern, and compared the resulting rearrangement metrics. The contested patterns of temporal evolution were:

- Chromothripsis followed by duplication events
- Whole chromosome duplication, followed by chromothripsis on both duplicated copies

The predictions from the rules are contrasted between the two scenarios in Supplementary Table 4 and the simulated rearrangement metrics are shown in Extended Data Figure 5.

Supplementary Table 4: Predictions based on the tenets in section D under two different rearrangement scenarios that give rise to a chromothripsis-like rearrangement pattern associated with three CN states.

Chromothripsis then two duplications	Whole chromosome duplication then chromothripsis involving both duplicated chromosomes
Tenet 2 predicts that some chromothripsis associated CN steps will be of size >1.	Tenet 1 predicts that all CN steps will be size 1, since no CN steps are being duplicated.
More than 3 CN states possible.	Only three CN states expected, since chromothripsis assumed to be a non-amplifying process.
Tenet 6 predicts that all chromothripsis associated CN jumps are of size 0.	Tenet 6 does not apply since after the first duplicate chromosome has undergone chromothripsis, the second duplicate cannot be thought as unrearranged anymore. Therefore CN jumps of size >0 are possible.

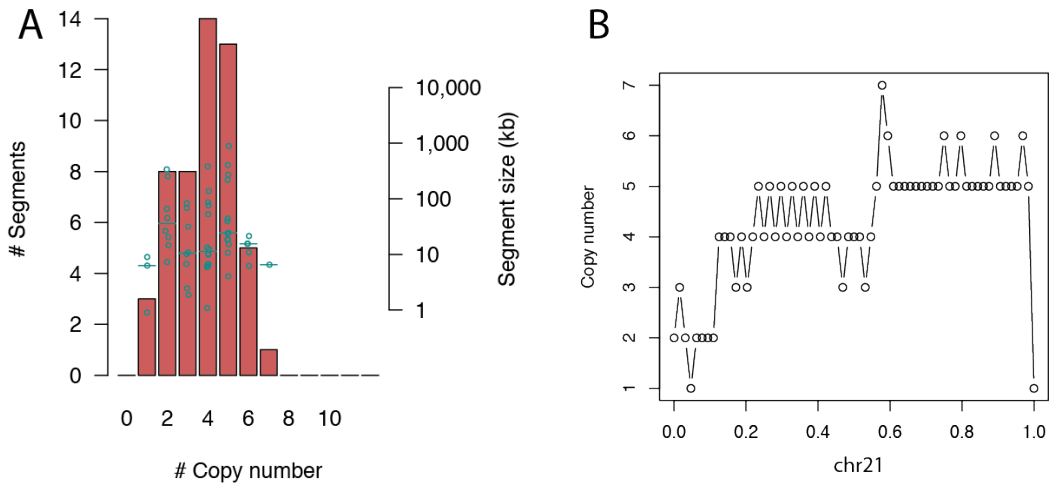
As expected, in the scenario with tandem duplications, occasionally duplication events overlap with each other and generate copy number states higher than three. In contrast, in the scenario where two independent chromothripsis events occur on two chromosome copies, copy number states are strictly confined to three states. In the scenario with tandem duplications, CN jump sizes associated with chromothripsis remained at 0 but many CN steps were of size >1. These metrics were reversed in the second scenario where only CN step size 1 was obtained but many rearrangement joins were associated with CN jumps between two different copy number states. In conclusion, the presented rearrangement metrics are sufficient to distinguish between the two alternative scenarios (Extended Data Figure 5).

F. Rearrangement history of the sequenced ALL samples

Sporadic iAMP21 ALL

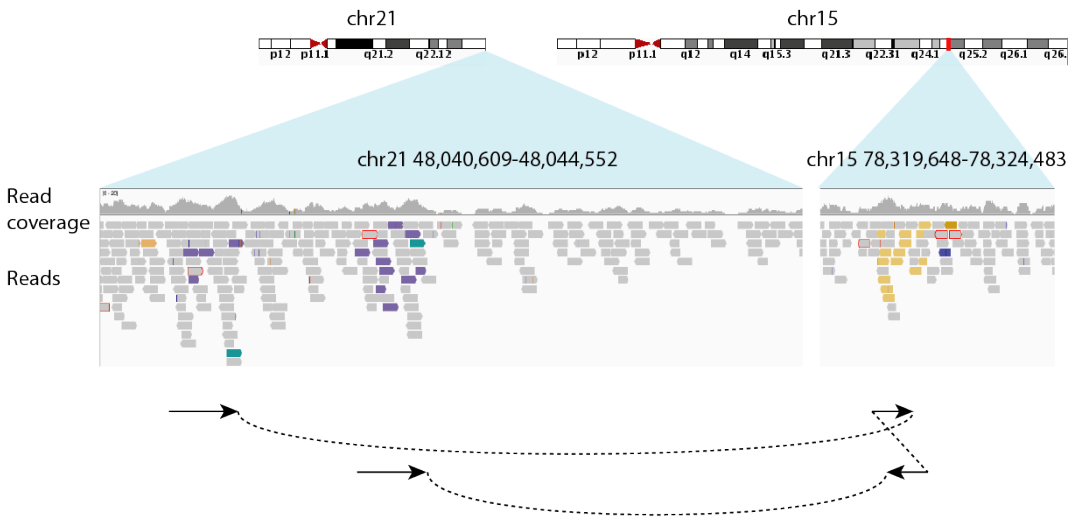
PD9020a

The copy number and rearrangement profile of PD9020a chromosome 21 is shown in Figure 1 and Supplementary Figure 10. CN steps associated with chromothripsis breakpoints are of size 1, and large fraction of CN jumps associated with chromothripsis breakpoints are of size >0 (Figure 1A-C, Supplementary Figure 10B). Together these indicate that chromothripsis occurred on an already amplified chromosome (tenets 1 and 6, scenario I). In contrast, if chromothripsis occurred prior to two cycles of BFB, many chromothripsis associated CN steps would be of size >1 (tenet 2, Extended Data Figure 4), but chromothripsis breakpoint associated CN jumps would be 0 (tenet 6, Extended Data Figure 4). Finally, the broad copy number pattern increases stepwise until a sharp CN drop demarcated by a fold-back rearrangement (Figure 1A, Supplementary Figure 10B). This is predicted if BFB occurred on an unrearranged chromosome (tenet 7), but would be unexpected if chromothripsis rearranged the chromosome prior to two cycles of BFBs (copy number trajectories on Extended Data Figure 4).



Supplementary Figure 10: Rearrangement metrics of patient PD9020a. A: Copy number state distribution of patient PD9020a chromosome 21. The frequency distribution is shown as a histogram and the size of each segment are shown as small circles. Small horizontal lines show the median segment size of each copy number. B: Copy number trajectory of patient PD9020a chromosome 21.

The q-telomeric DNA double-stranded break initiating the first BFB cycle is repaired by incorporating a shard of ~500bp derived from chromosome 15 (Supplementary Figure 11). The shard has been removed in Figure 1A, and the rearrangement is instead shown as a simple fold-back rearrangement for illustration purposes.



Supplementary Figure 11: An IGV diagram of the shard on chromosome 15 involved in the fusion step of the first BFB cycle.

In conclusion, rearrangements in patient PD9020a iAMP21 were initiated by two BFB cycles, with the first DSB near the q-telomere and the second at ~18Mb. A chromothripsis event then completed the sequence of rearrangements (Figure 1E).

PD9021a

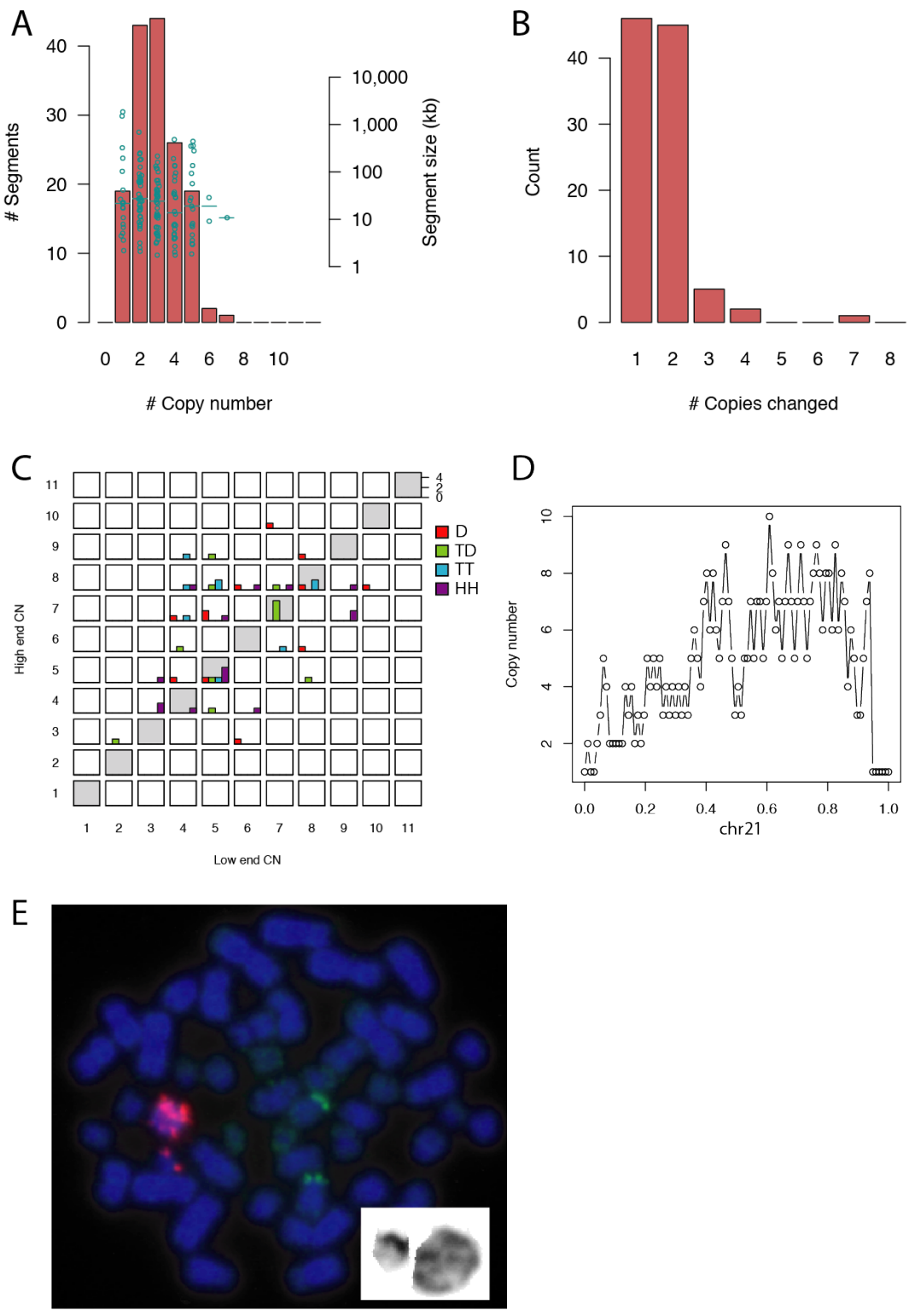
The chromosome 21 copy number and rearrangement profile of PD9021a is shown in Figure 3, with the associated rearrangement metrics; FISH and partial karyotype, shown in Supplementary Figure 12.

A step-wise increasing copy number pattern and a foldback rearrangement, demarcating a steep telomeric copy number drop, indicate an initiating BFB event (tenet 7, Figure 3A, Supplementary Figure 12D). Chromothripsis-associated rearrangements, which constitute the majority of the chromosome 21 rearrangements in this sample, are associated with numerous CN jumps of size >0 (Supplementary Figure 12C), indicating that chromothripsis did not occur on an unrearranged chromosome (tenet 6). Altogether, rearrangement metrics are consistent with the model that chromothripsis occurred on a chromosome first amplified by BFB cycles.

At both around 28Mb and 42Mb, there are four copy number states associated with chromothripsis rearrangements, indicating that chromothripsis was associated with up to four different copy number states (Figure 3A, Supplementary Figure 12D). This is consistent with the model that two BFB cycles occurred before chromothripsis, generating up to four copies of the amplified region of the derivative chromosome 21. The BFB breakpoint associated with the second BFB cycle occurred probably at ~22-25Mb (Figure 3A), and was later obscured by the cluster of chromothripsis rearrangements in this region. Supplementary Figure 13 shows a schematic example of how this may occur.

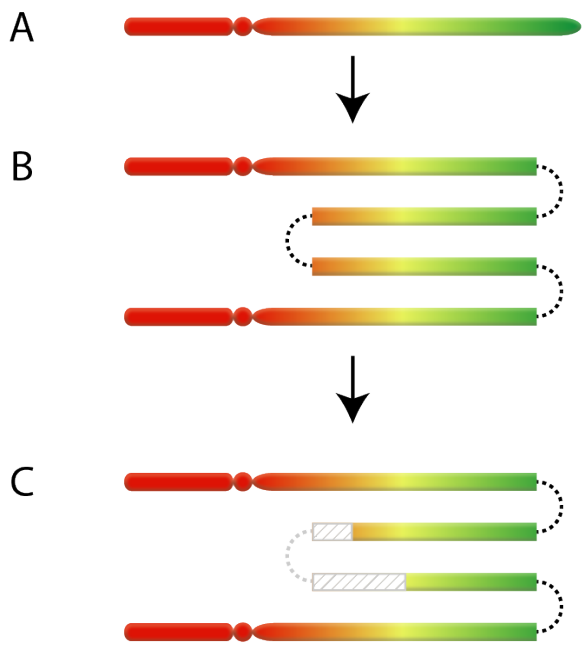
The copy number of chromosome 21 is around 4 proximal to ~25Mb, but increases to around 7-8 from ~25Mb towards q-telomere (Figure 3A, Supplementary Figure 12D). Moreover, CN steps proximal to 25Mb are of size 1, whereas most of them distal to 25Mb, constituting around half of all CN steps on chromosome 21, are of size 2 (Supplementary Figure 12B,D). This suggests that the region from ~25Mb to the first BFB breakpoint underwent a late duplication event, doubling the CN steps generated by chromothripsis (tenet 2). The duplication event is supported by cytogenetic analysis, which showed a ring chromosome 21 (Supplementary Figure 12E).

In summary, rearrangements of chromosome 21 in patient PD9021a likely started with two BFB cycles. The derivative chromosome then underwent chromothripsis. Finally, the major component of the derivative chromosome was duplicated, probably in association with the ring chromosome formation. Due to the complexity generated by chromothripsis, we could not pinpoint any simple deletion or duplication events and separate them from chromothripsis associated copy number changes.



Supplementary Figure 12: Rearrangement metrics of patient PD9021a chromosome 21. Copy number state distribution (A), copy number step distribution (B), copy number jump distribution (C) and copy number trajectory (D). Metaphase spread hybridized with the *TEL/AML1* ES probe (Abbott Molecular) which shows multiple signals for *RUNX1* (red) clustered on a single abnormal chromosome. The

discrete pair of red signals indicates *RUNX1* on the normal chromosome 21. The two pairs of green signals indicate the location of *ETV6* on the normal chromosomes 12. Inset shows a partial G-banded karyotype of the chromosomes 21 of this patient: the abnormal chromosome 21 is the large ring chromosome on the right compared to the normal chromosome 21 on the left. (E). In A, the frequency distribution is shown as a histogram and the size of each segment are shown as small circles. Small horizontal lines show the median segment size of each copy number. In C, D – deletion type rearrangement link, TD – tandem duplication type, TT – tail to tail rearrangement, HH – head to head rearrangement.



Supplementary Figure 13: Schematic example of how earlier rearrangement joins can be removed by later rearrangements. A chromosome undergoes two cycles of BFB (A to B), generating two fold-back rearrangements (dashed arcs). A subsequent deletion overlapping a fold-back rearrangement (C, grey rectangles) then deletes one of the previously generated fold-back rearrangements (grey dashed arc).

PD9022a

Copy number and rearrangement patterns for chromosome 21 of PD9022a are shown in Extended Data Figure 6. The associated rearrangement metrics are shown in Supplementary Figure 14. Cytogenetic data were not available for this patient, who showed a normal karyotype. SNP6.0 data showed that regions from p-telomere to ~20Mb and from ~45Mb to q-telomere were heterozygous, with DNA present from one copy of both the iAMP21 and the wild-type chromosome 21 (Supplementary Figure 15).

This patient did not exhibit the CN and fold-back rearrangement patterns seen in other sporadic iAMP21 patients in this study. The copy number of chromosome 21 did not increase stepwise towards q-telomere (Extended Data Figure 6, Supplementary Figure 14D) and the q-telomeric region was not lost from the iAMP21 chromosome (Extended Data Figure 6, Supplementary Figure 14D). Therefore this iAMP21 chromosome was not initiated by a q-telomeric DSB and ensuing BFB cycles.

However, we do see a foldback rearrangement at ~25Mb, demarcating a drop in copy number from 4 to 0 on the iAMP21 chromosome (Extended Data Figure 6, Supplementary Figure 14D), and a second potential fold-back-like rearrangement between ~42Mb and 44Mb (curved arrow in Extended Data Figure 6). The latter may have been a deleted fold-back rearrangement, such as the one schematically shown in Supplementary Figure 13. Therefore it is possible that two BFB cycles did occur on this chromosome, although not through losing the q-telomeric region. One possible temporal order of events is that the fold-back rearrangement at ~25Mb initiated the first BFB cycle, as the region immediately upstream from it is lost. The fold-back-like rearrangement between 42Mb and 44Mb would then be the second breakage-fusion breakpoint.

If this model is true, how would the region from p-telomere to ~20Mb be preserved in the iAMP21 chromosome? No read pairs were visible spanning the copy number breakpoint at ~20Mb. However, when searching for unmapped reads whose mates mapped to ~20Mb, we detected unmapped split reads linking this breakpoint to a region at ~24.4Mb (Extended Data Figure 6). Therefore it seems that prior to or during the BFB cycles, the region from p-telomere to ~20Mb became joined to 24.4Mb, explaining how this region was retained even if the fold-back rearrangement at ~25Mb were the initiating BFB breakpoint. The rearrangement between ~20Mb and ~24.4Mb would also have been important in recruiting a centromere to the iAMP21 chromosome (Extended Data Figure 6).

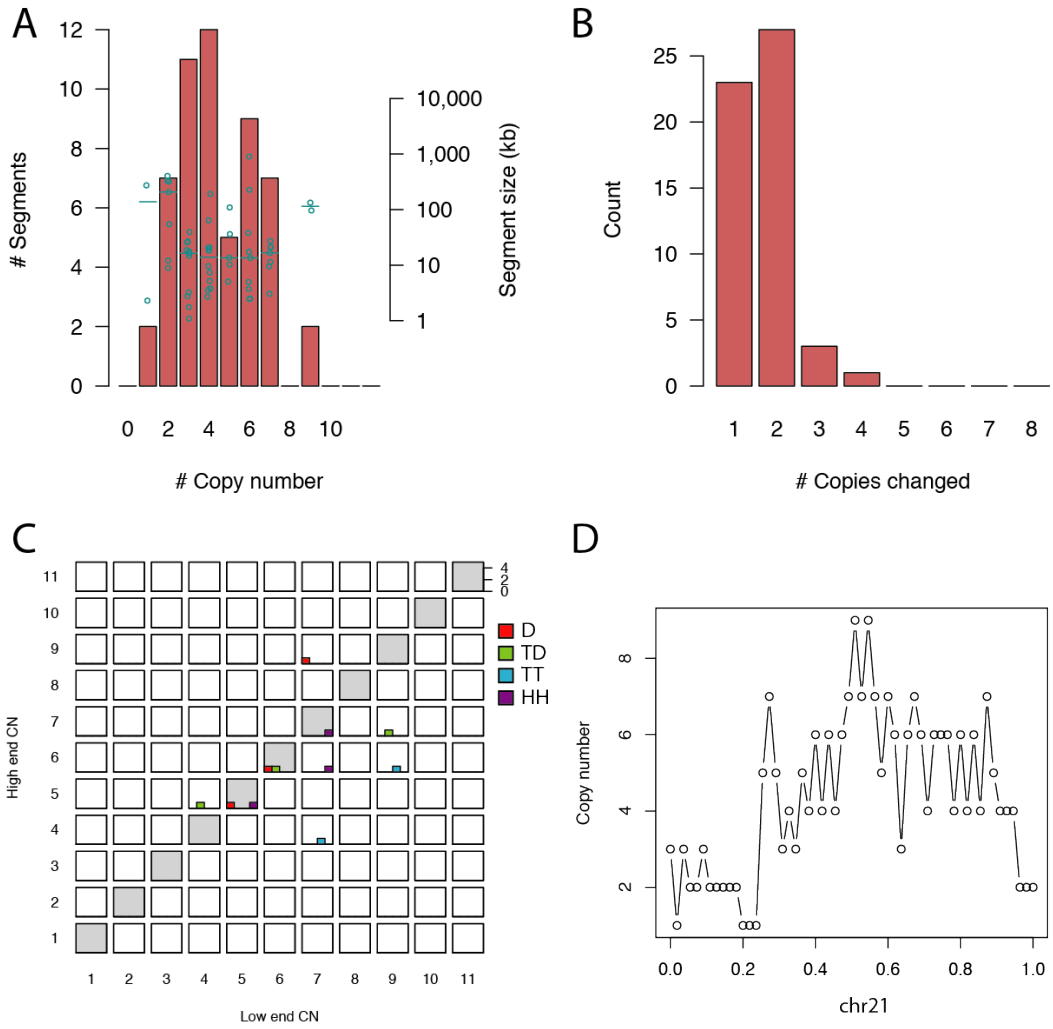
An obvious jagged copy number pattern is present in the amplified section of the chromosome (Extended Data Figure 6, Supplementary Figure 14D). Although extracting CN jumps failed in many cases due to extremely close clustering of the chromothripsis associated breakpoints, half of the resolved CN jumps had size >0 (Supplementary Figure 14C), indicating chromothripsis occurred after the amplification had commenced (tenet 6).

Most of the chromothripsis CN steps were of size 2 (Supplementary Figure 14D), suggesting that a large-scale duplication occurred after chromothripsis, duplicating almost the entire derivative chromosome (tenet 2). This model is consistent with the overall copy number state distribution of the chromosome, with copy number states

of up to 8 copies of some regions of the iAMP21 chromosome (Extended Data Figure 6, Supplementary Figure 14D). Two initial BFB cycles would have generated up to 4 copies of the amplified region of the iAMP21 chromosome, and a large-scale duplication would have taken the highest copy number from 4 to 8.

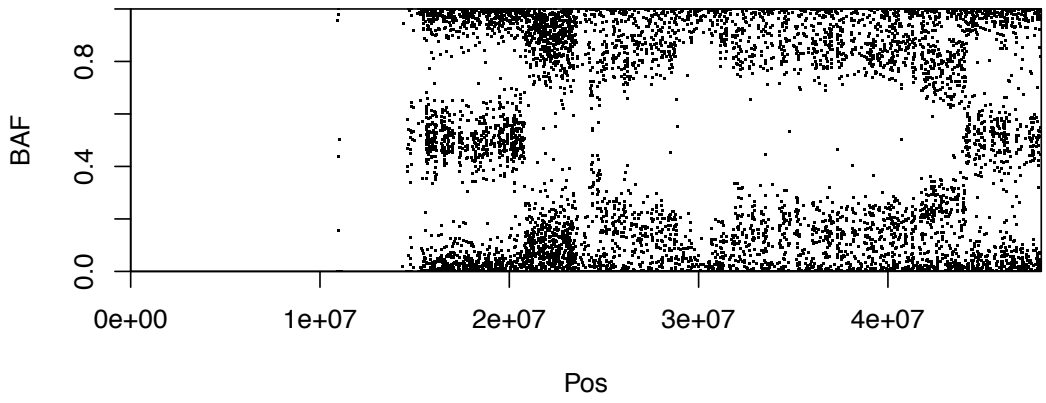
The final iAMP21 chromosome has one copy of the q-telomeric region (Extended Data Figure 6, Supplementary Figure 15), indicating that the final duplication event did not encompass the entire iAMP21 chromosome. This observation along with copy number alterations after the late duplication event would explain why intermediate copy number states remained (1, 3, 5 and 7) after the large-scale duplication event.

In summary, the nature of the initiating rearrangement in this case is not completely clear. There is however strong support for a sequence of events where chromothripsis occurred on a chromosome amplified by two cycles of BFB, and was followed by a large-scale duplication event overlapping with most of the chromothripsis associated breakpoints.



Supplementary Figure 14: Rearrangement metrics for patient PD9022a chromosome 21. Copy number state distribution (A), copy number step distribution

(B), copy number jump distribution (C), copy number trajectory (D). In A, the frequency distribution is shown as a histogram and the size of each segment is shown as small circles. Small horizontal lines show the median segment size of each copy number. In C, D – deletion type rearrangement link, TD – tandem duplication type, TT – tail to tail rearrangement, HH – head to head rearrangement.



Supplementary Figure 15: B-allele frequency of chromosome 21 from SNP6.0 data of patient PD9022a.

PD9023a

The copy number and rearrangement pattern of chromosome 21 from patient PD9023a is shown in Figure 3. The associated rearrangement metrics and cytogenetic features are shown in Supplementary Figure 16.

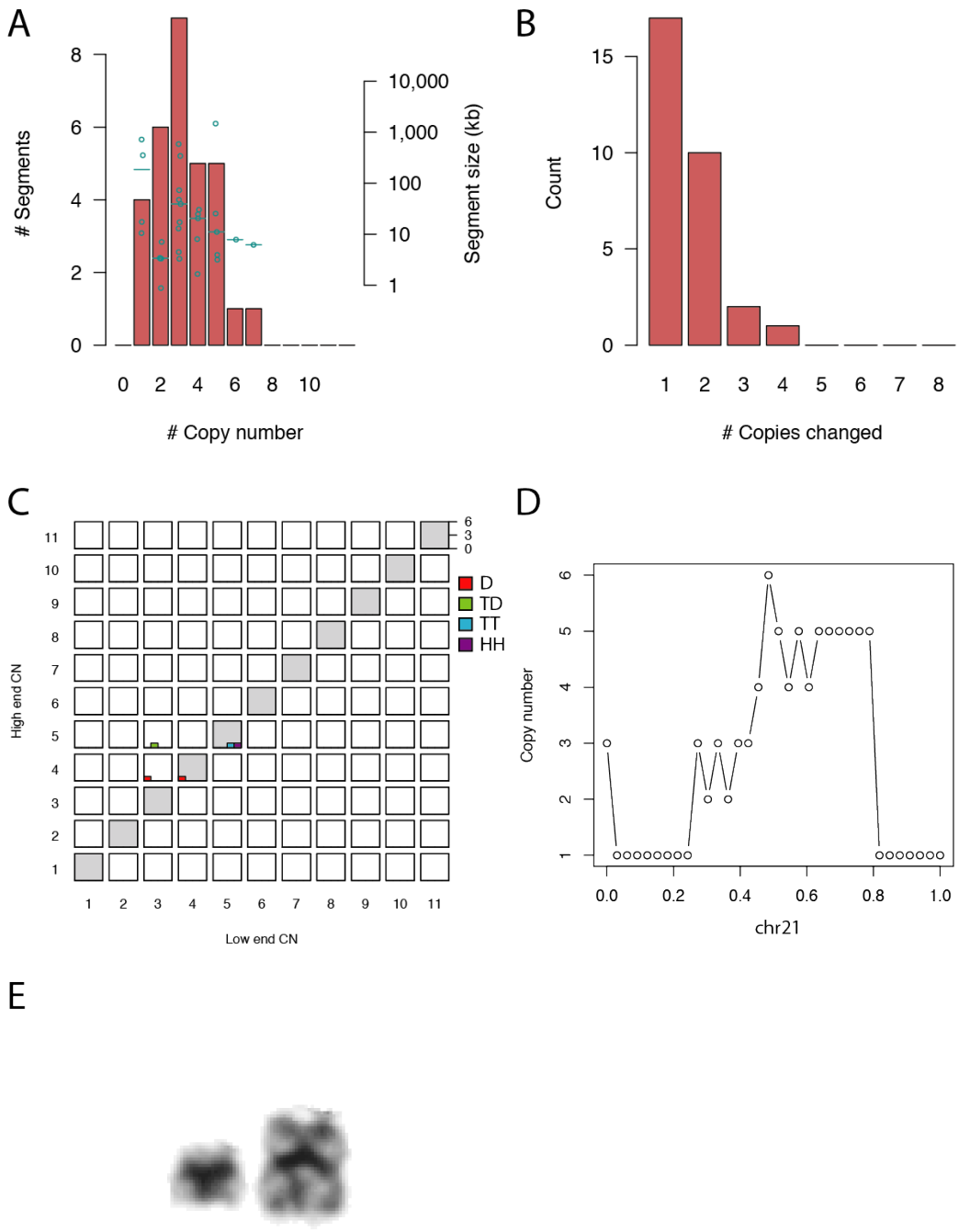
This patient has a copy number pattern indicative of two BFB cycles, with first breakage-fusion breakpoint at ~45Mb and the second one at ~22Mb (Supplementary Figure 17). There are few copy number changes at ~23Mb and ~30Mb. CN steps at these regions are almost exclusively -1 or +1 (Supplementary Figure 16D). This indicates that the corresponding rearrangements occurred after the two BFB cycles, as otherwise they would have been associated with CN step sizes >1 (tenets 1 and 2). There are insufficient rearrangements in these regions to meet criteria for chromothripsis.

The cytogenetic pattern of the iAMP21 chromosome indicates an iso-chromosome (Supplementary Figure 16E). This pattern is consistent with the model in which the second BFB step was the final major rearrangement of this chromosome.

A (PCR validated) head-to-tail rearrangement between ~10Mb and ~30Mb (Figure 3) is consistent with the model that the chromosome 21 centromere is linked to the iso-chromosome.

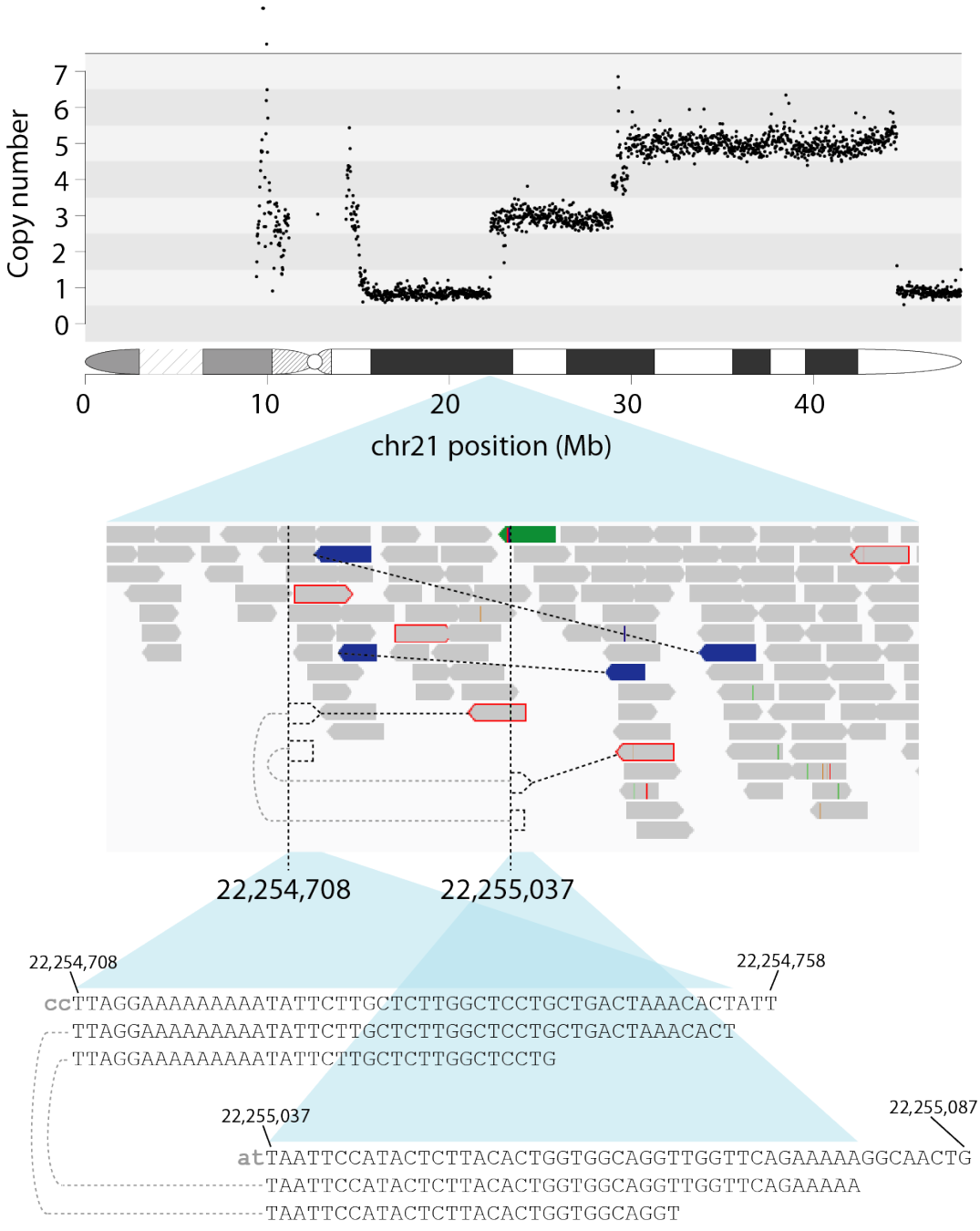
In summary, two cycles of BFB, first at ~45Mb and second at ~22Mb were the initiating rearrangements. They were followed by a series of rearrangements altering

the copy numbers at around ~23Mb and ~30Mb and linking the centromere to the resultant iAMP21 chromosome.



Supplementary Figure 16: Rearrangement metrics of chromosome 21 in patient PD9023a. Copy number state distribution (A), copy number step distribution (B), copy number jump distribution (C), copy number trajectory (D). Partial karyotype showing the normal chromosome 21 on the left and the iAMP21 chromosome on the right, which is in the form of a large metacentric chromosome, likely isochromosome (E). In A, the frequency distribution is shown as a histogram and the

size of each segment is shown as small circles. Small horizontal lines show the median segment size of each copy number. In C, D – deletion type rearrangement link, TD – tandem duplication type, TT – tail to tail rearrangement, HH – head to head rearrangement.



Supplementary Figure 17: Fold-back rearrangement in patient PD9023a. Two read pairs and two split reads supporting the fold-back rearrangement at ~22Mb region.

Dotted black lines connect paired reads and dotted grey lines connect split reads. The two connected breakpoints are shown by vertical dotted lines.

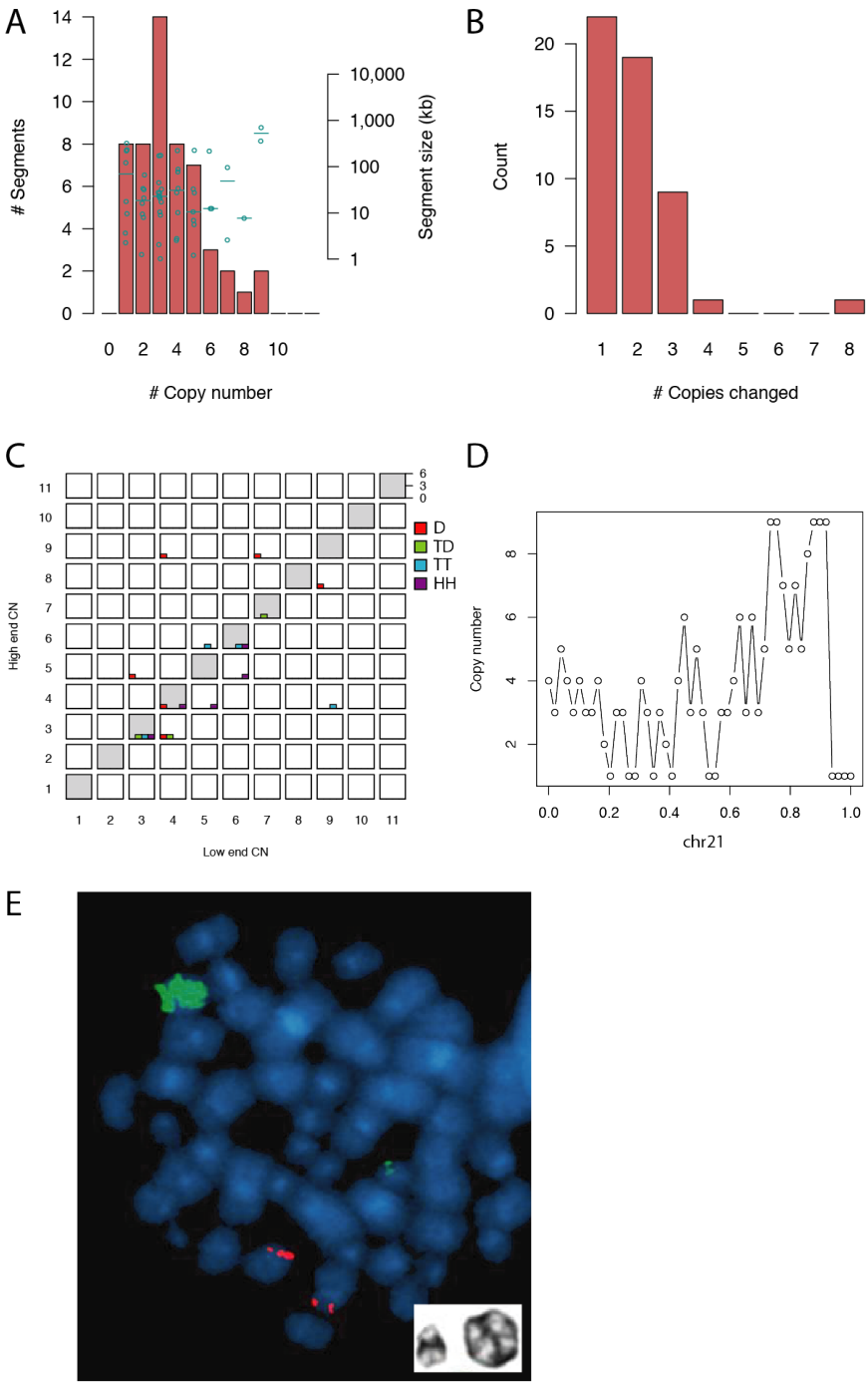
PD4117a

The copy number and rearrangement pattern of chromosome 21 in patient PD4117a is shown in Figure 3. The associated rearrangement metrics and FISH and cytogenetic features are shown in Supplementary Figure 18.

Copy number increases stepwise towards ~45Mb where a fold-back rearrangement demarcates a steep copy number drop (Figure 3, Supplementary Figure 18D). Within the amplified region there is a jagged copy number pattern with numerous back-and-forth rearrangements indicative of chromothripsis (Figure 3, Supplementary Figure 18D). Half of the CN jumps join differing copy numbers (Supplementary Figure 18C). A substantial fraction of the chromothripsis-associated CN steps are of size >1. However, the CN step sizes are uniform throughout the chromosome. If these CN steps were amplified by BFB cycles following chromothripsis, the region distal to ~34Mb, which has a much higher copy number than the region proximal to ~34Mb, should have larger CN step sizes. Therefore it is more likely that chromothripsis occurred after the BFB cycles, which was followed by large-scale duplication. Together, the rearrangement metrics support a model in which the chromosome was first amplified by BFB and subsequently underwent chromothripsis (tenet 6, tenet 7).

The maximal copy number is 5 and 9 proximal and distal to ~34Mb, respectively (Figure 3). This copy number pattern is consistent with the model in which two BFB cycles generate a derivative chromosome with a maximum of 2 and 4 copies proximal and distal to ~34Mb occurring first. Following the BFB cycles and chromothripsis, a whole-chromosome duplication takes the maximal copy number to 4 and 8 on the iAMP21 chromosome proximal and distal to ~34Mb, respectively. The second presumed BFB breakpoint, probably occurred somewhere in the ~30-35Mb region (Figure 3), but this event was not observed in this sequencing screen for unknown reasons. Cytogenetic analysis showed the iAMP21 chromosome to be in the form of a ring (Supplementary Figure 18E). This is consistent with a large-scale duplication occurring as the last major event following chromothripsis.

In summary, we observed the first cycle of BFB that initiated the rearrangements, and presumed that a second cycle followed for which we were unable to map the rearrangement. This event was followed by chromothripsis and finally by a large-scale duplication associated with ring chromosome formation.



Supplementary Figure 18: Rearrangement metrics of chromosome 21 in patient PD4117a. Copy number state distribution (A), copy number step distribution (B), copy number jump distribution (C), copy number trajectory (D). Metaphase spread hybridized with the TEL/AML1 DC DF probe (Cytocell) shows multiple signals for *RUNX1* (green) clustered on a single abnormal chromosome. The discrete pair of green signals indicates *RUNX1* on the normal chromosome 21. The two pairs of red signals indicate the location of *ETV6* on the normal chromosomes 12. The chromosomes are counterstained with DAPI (blue). Inset shows a partial G-banded karyotype of the chromosomes 21 of this patient: the abnormal chromosome 21 is the large ring chromosome on the right compared to the normal chromosome 21 on

the left (E). In A, the frequency distribution is shown as a histogram and the size of each segment are shown as small circles. Small horizontal lines depict median segment size of given copy number. Small horizontal lines show the median segment size of each copy number. In C, D – deletion type rearrangement link, TD – tandem duplication type, TT – tail to tail rearrangement, HH – head to head rearrangement.

der(15;21) iAMP21 ALL associated with rob(15;21)c

PD7170a

The copy number and rearrangement pattern of chromosomes 15 and 21 of the der(15;21) iAMP21 of patient PD7170a are shown in Figure 2. The associated rearrangement metrics and cytogenetic features are shown in Figure 2 and Supplementary Figure 19.

The major CN states on chromosomes 15 and 21 are 1, 3 and 5 (Figure 2A, Supplementary Figure 19), indicating major copy numbers of 0, 2 and 4 on the der(15;21) chromosome. The major CN step size is 2 over the chromosome (Figure 2A-B). Together these observations indicate whole-chromosome duplication as the final rearrangement event, duplicating the CN step sizes from 1 to 2 (tenet 2). Cytogenetic analysis shows an isochromosome to be the result of this duplication.

Prior to this duplication event, a large number of back-and-forth rearrangements demarcate numerous copy number changes that alternate between three copy number states (Figure 2). This pattern is highly suggestive of chromothripsis, but involves three CN states instead of the expected two (Supplementary Figure 2B).

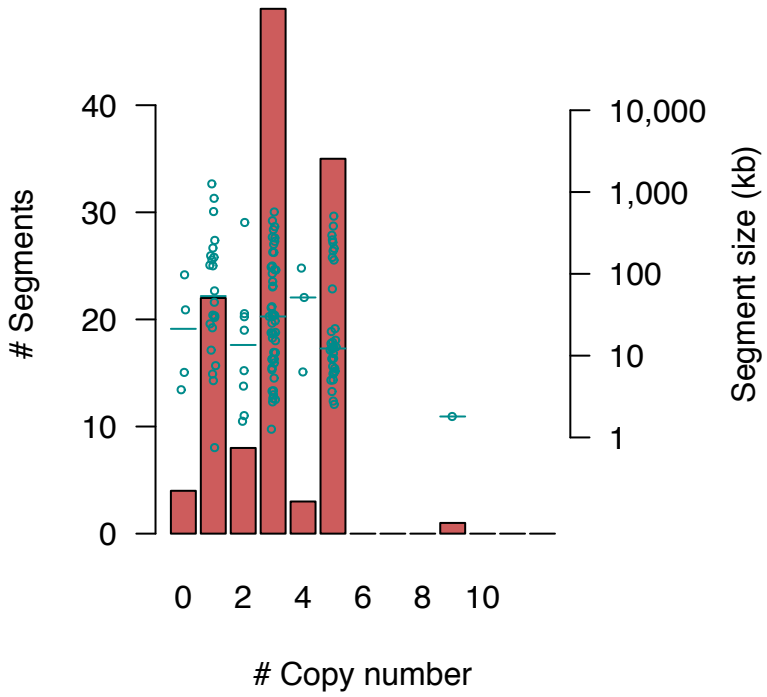
Two simple sequences of rearrangements could generate a chromothripsis-like rearrangement pattern involving three CN states. In the first scenario, chromothripsis could occur on an unrearranged chromosome, generating oscillations across two CN states. This might be followed by partial duplication, taking some retained segments into higher copy number, while preserving the rearrangement pattern. In the second scenario, chromothripsis could occur simultaneously on two copies of an unrearranged chromosome (both sister chromatids post-replication).

As described in section D, and validated through simulations in section E the former scenario is associated with CN step sizes >1 and CN jumps of size 0 onto chromothripsis breakpoints. In contrast, the latter scenario is associated with CN step size 1, CN jump sizes >0, and a maximum total CN steps of 2 on the derivative chromosomes (Supplementary Table 4, Extended Data Figure 5).

Der(15;21) of PD7170a had almost exclusively CN steps of size 2 which would have been size 1 prior to whole-chromosome duplication. In contrast, there were very few CN breakpoints with CN step size 4 (CN step size 2 prior to the duplication). There were many CN jumps of size >0, and a large number of segments with copy number 4 (on the der(15;21)) but very few with copy number >4. In conclusion, the rearrangement metrics strongly favor the model in which the chromothripsis pattern associated with 3 copy number states was acquired through chromothripsis events on two copies of rob(15;21)c (Supplementary Table 4, Extended Data Figure 5).

Two lines of evidence support a model in which the two duplicated rob(15;21)c copies that underwent chromothripsis in fact shattered in the same event and were repaired into a single der(15;21) chromosome. First, cytogenetic analysis showed a single isochromosome, and thus the progenitor of this isochromosome was a single der(15;21) chromosome. Second, we found three fold-back-like rearrangements (Supplementary Figure 4D, Extended Data Figure 7) of the der(15;21) chromosome. A fold-back-like rearrangement joins two segments comprising DNA from the same chromosomal region (Supplementary Figure 4D, Extended Data Figure 7). When a fold-back-like rearrangement is formed, two copies of the same chromosomal region must be present and available for rearrangements through nearby double strand DNA breaks. If two der(15;21) sister chromatids underwent chromothripsis together, this would indeed be the case, as there would be two copies of the der(15;21) shattering and being repaired at the same time. On the other hand, if two copies of der(15;21) underwent chromothripsis independently, fold-back-like rearrangements could not occur, as only one copy of each chromosomal location is present among fragments generated in either chromothripsis event (Extended Data Figure 8).

In summary, the initial rearrangement was a chromothripsis event affecting two sister chromatids of der(15;21) simultaneously. Segments of the two sister chromatids were repaired into one derivative chromosome, which was later duplicated through iso-chromosome formation.



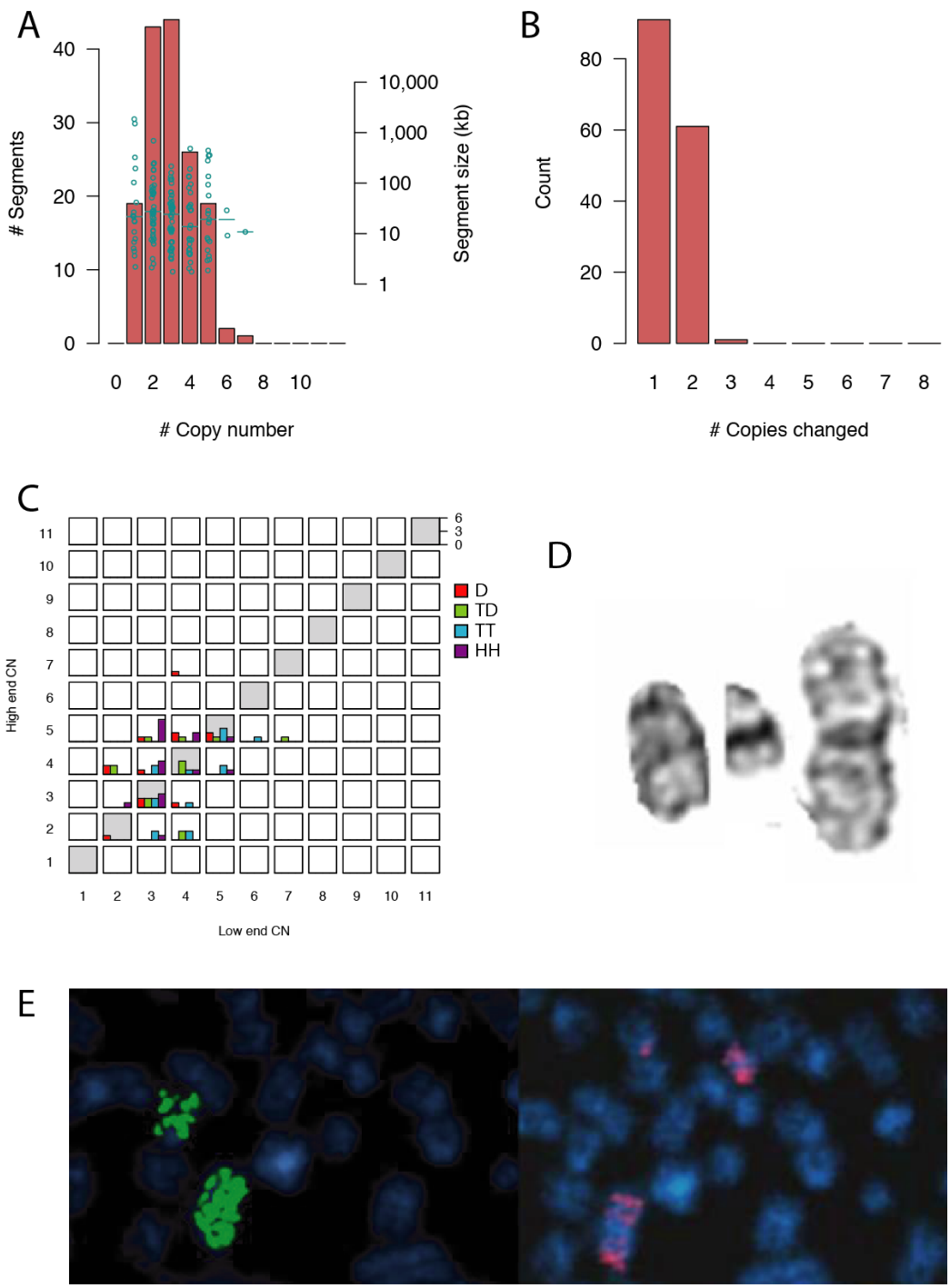
Supplementary Figure 19: Copy number state distribution of patient PD7170a chromosomes 15 and 21. The frequency distribution is shown as a histogram and the size of each segment are shown as small circles. Small horizontal lines depict median segment size of given copy number.

PD7171a

The copy number and rearrangement pattern of PD7171a is shown in Figure 3. The associated rearrangement metrics and cytogenetic features are shown in Supplementary Figure 20.

The rearrangement and copy number patterns of the der(15;21) chromosome show a large number of back-and-forth rearrangements in all four possible orientations connecting copy number segments of few copy number states (Figure 3). This is indicative of chromothripsis. Due to the complexity of the rearrangement pattern and CN states, it appears that chromothripsis was not the only event shaping the chromosome, but we cannot resolve the exact temporal evolution in this patient. However, we presume the der(15;21) chromosome endured a late and extensive partial chromosome duplication, since almost half of the CN steps are of size 2 (Supplementary Figure 20B, tenet 2).

In summary, due to the complexity of the rearrangement pattern, a clear view of the temporal evolution of rearrangements could not be obtained in this case. However, rearrangement metrics support the model in which the der(15;21) chromosome underwent chromothripsis early, probably first, which was followed by a duplication amplifying the CN steps of almost half of the chromothripsis breakpoints.



Supplementary Figure 20: Rearrangement metrics of chromosomes 15 and 21 for patient PD7171a. Copy number state distribution (A), copy number step distribution (B), copy number jump distribution (C). D shows a partial G-banded karyotype of the chromosomes 15 and 21 of this patient: from left to right are the normal chromosome 15, the normal chromosome 21 and the abnormal der(15;21) in the form of a large metacentric chromosome, likely an isochromosome. In E, the two side by side images represent sequential FISH on the same metaphase using whole chromosome paints specific for chromosomes 15 (WCP15) (left) and 21 (WCP21) (right). In the left image the left hand and right hand green chromosomes indicate the normal and the der(15;21), respectively. In the right hand image, the top red

signal indicates the normal chromosome 21 and the lower chromosome, which also painted green in the left hand image, shows stripes of red indicating that regions of chromosome 21 are interspersed with chromosome 15 within this iAMP21 chromosome, confirming that the abnormal chromosome is derived from both chromosomes 15 and 21. The chromosomes are counterstained with DAPI (blue). In A, the frequency distribution is shown as a histogram and the size of each segment is shown as small circles. Small horizontal lines depict median segment size of given copy number. In C, D – deletion type rearrangement link, TD – tandem duplication type, TT – tail to tail rearrangement, HH – head to head rearrangement.

PD10008a

The copy number and rearrangement pattern of chromosomes 15 and 21 in patient PD10008a is shown in Figure 3. The associated rearrangement metrics and cytogenetic features are shown in Supplementary Figure 21.

The rearrangement pattern is indicative of chromothripsis, with back-and-forth rearrangements in all four orientations joining alternating copy number segments.

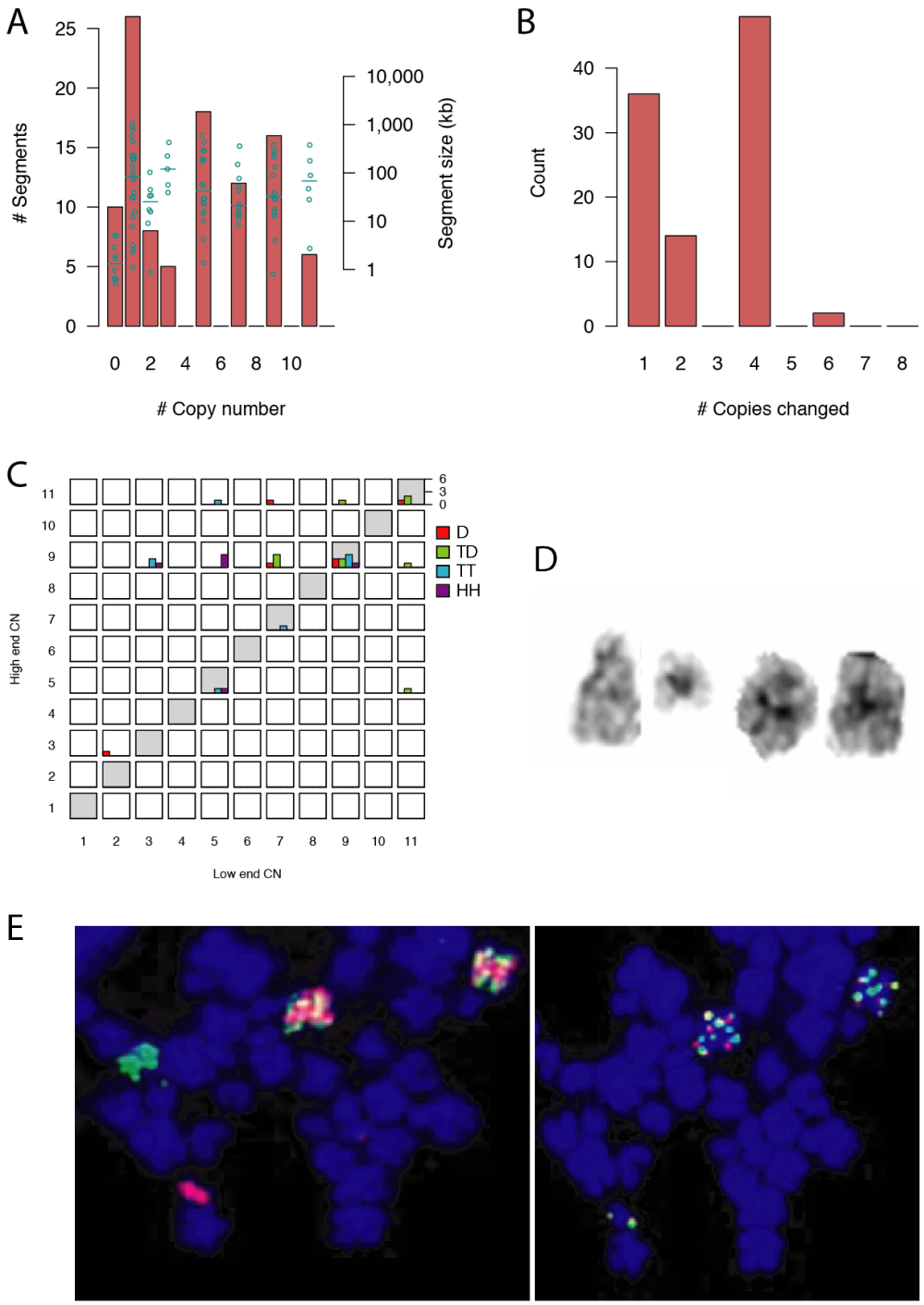
The major copy number states formed by the der(15;21) chromosome are 0, 4 and 8, with some copy number states at 2, 6 and 10 (Figure 3, Supplementary Figure 21A). Similarly, there are few CN steps of size 1 or 2, none at step size 3 with the most frequent CN step size being 4 (Supplementary Figure 21B). This CN state and step distribution suggests that two chromosome duplication events followed chromothripsis, giving rise to 4 as the most frequent CN step size (tenet 2). The model is supported by the cytogenetic data, which indicate the presence of two copies of a der(15;21) isochromosome. Therefore, the derivative chromosome that arose from chromothripsis first duplicated through isochromosome formation, then the isochromosome was duplicated through an aberrant mitosis.

Chromothripsis was mainly associated with 3 (total) copy number states on the der(15;21) chromosomes, 0, 4 and 8 (Figure 3). These correspond to copy numbers 0, 1 and 2 prior to the 3 duplication events. We found a fold-back-like rearrangement within the der(15;21) chromosome (Supplementary Figure 22). Following the reasoning applied to PD7170a, therefore, there is strong support for simultaneous chromothripsis on two der(15;21) sister chromatids as the initiating rearrangement event in this case.

We observed multiple copy number segments at intermediate (total) copy numbers 2, 6 and 10 on the der(15;21) chromosome (Figure 3, Supplementary Figure 21A), likely due to duplications or deletions occurring at the initial isochromosome stage but before the isochromosome duplicated. During the single isochromosome stage, the major copy numbers of the der(15;21) chromosome were 0, 2 and 4. Deletions or duplications generating copy numbers 1, 3 and 5 at this stage would be duplicated into copy numbers 2, 6 and 10, respectively, during the last whole-(iso)chromosome duplication event.

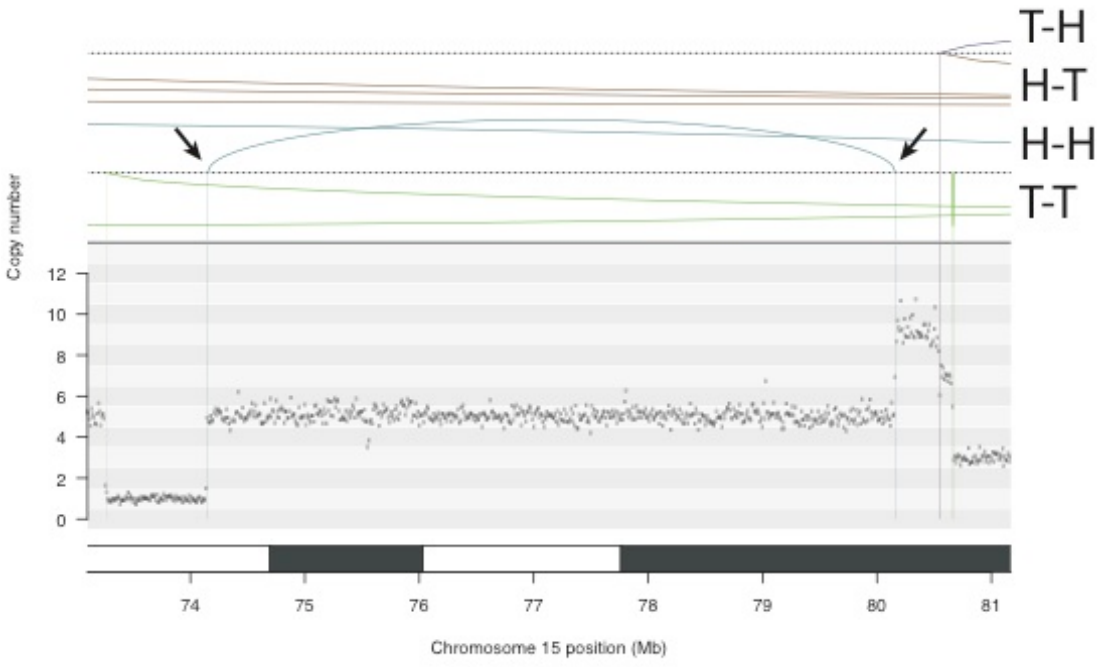
In summary, a simultaneous chromothripsis on two rob(15;21)c sister chromatids was the initiating event of the der(15;21) chromosome. This chromosome was then

duplicated through an isochromosome formation event. At this stage, several deletion and duplication events occurred before the final major event when the isochromosome duplicated. Rearrangements associated with CN step size 1 were generated after the isochromosome duplication.



Supplementary Figure 21: Rearrangement metrics of chromosomes 15 and 21 for patient PD10008a. Copy number state distribution (A), copy number step distribution (B), copy number jump distribution (C). D shows a partial G-banded karyotype of the chromosomes 15 and 21 of this patient: from left to right are the

normal chromosome 15, the normal chromosome 21 and two copies of abnormal rob(15;21) in the form of metacentric chromosomes, likely isochromosomes. In E, the two side by side images represent sequential FISH on the same metaphase. The left hand cell is simultaneously hybridized with whole chromosome paints specific for chromosomes 15 (WCP15, green) and 21 (WCP21, red). The green chromosome on the left and the red chromosome below indicate the normal chromosomes 15 and 21, respectively. On the right of the image the two chromosomes showing interspersed patterns of red and green indicate two copies of the der(15;21) with intermingling of chromosomes 15 and 21 material along their lengths. In the right hand image, the same cell is hybridized with two probes specific for *RUNX1* (exons 1-5, cosmid ICRF C0664 in red; exons 6-7, cosmid ICRF HO8116 in green) within the common region of amplification, one is labelled green and the other is labelled red. When closely apposed the fused red-green signal appears yellow. Note the paired yellow signals indicating the common region of amplification on the normal chromosome 21 at the bottom of the image. On the two copies of the der(15;21) iAMP21 chromosome there is a mixture of red, green and yellow signals along the length of the chromosomes indicating the complexity of their structural rearrangements. The similarity of these patterns between the two chromosomes supports the suggestion that the whole chromosome duplication followed the chromothripsis and other rearrangements. In A, the frequency distribution is shown as a histogram and the size of each segment is shown as small circles. Small horizontal lines depict median segment size of given copy number. In C, D – deletion type rearrangement link, TD – tandem duplication type, TT – tail to tail rearrangement, HH – head to head rearrangement.



Supplementary Figure 22: A fold-back-like rearrangement on chromosome 15 in patient PD10008a.

PD10009a

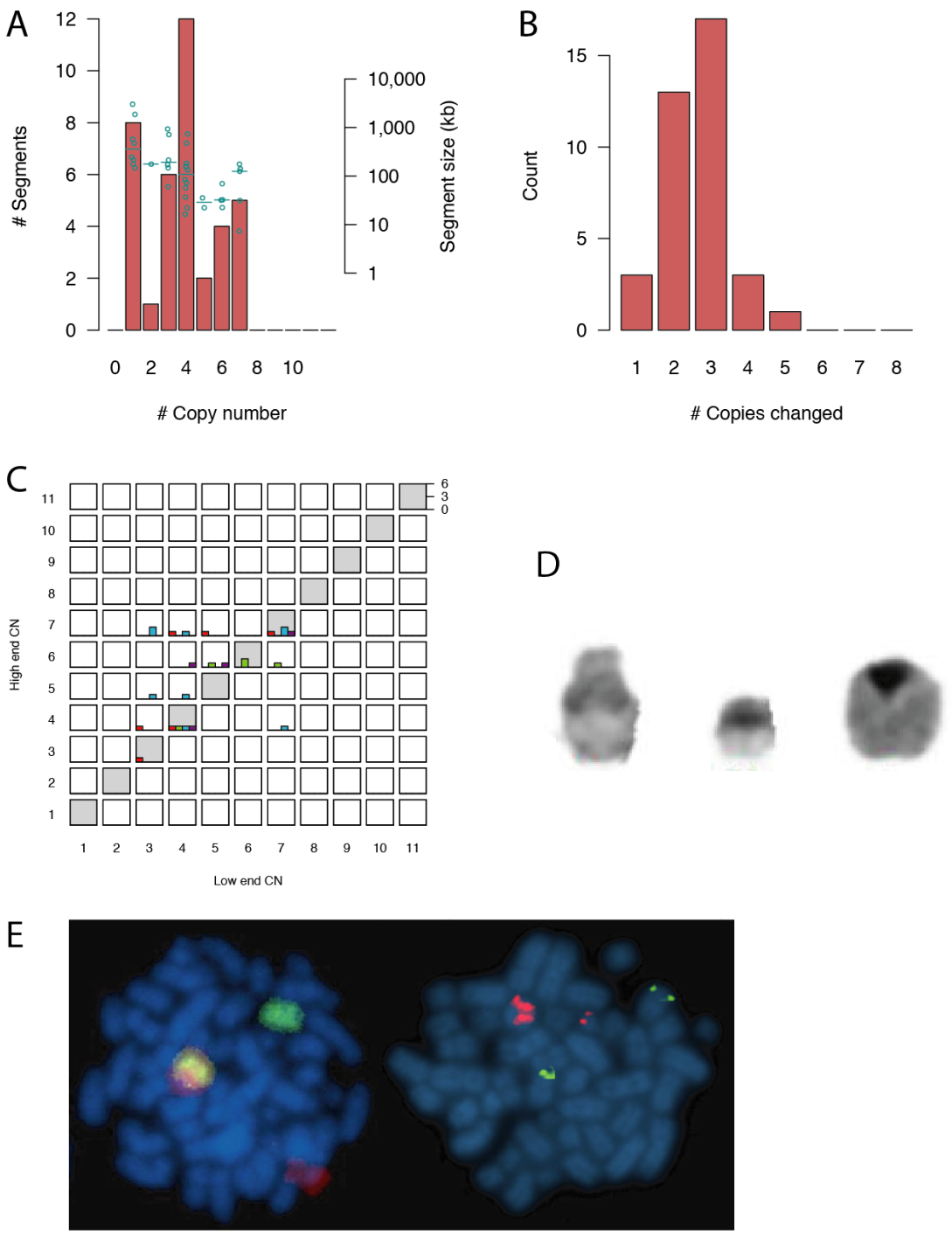
The copy number and rearrangement pattern of chromosomes 15 and 21 in patient PD10009a is shown in Figure 3. The associated rearrangement metrics and cytogenetic features are shown in Supplementary Figure 23.

The rearrangement metrics of chromosomes 15 and 21 of this sample are remarkably similar to those of PD7170a. Rearrangement and copy number patterns point to chromothripsis involving three copy number states (Figure 3B) and the copy number state distribution shows that there are few copy number states between the three main states (Supplementary Figure 23A). Moreover, there are only few CN steps of size 1 (Supplementary Figure 23B). Taken together, this indicates one or more whole-chromosome duplication events occurred after chromothripsis.

Cytogenetic analysis showed the der(15;21) to be in the form of a ring chromosome. Therefore the aggregated data was consistent with one or two rounds of whole-chromosome duplication associated with a ring chromosome following chromothripsis.

As in PD7170a, chromothripsis associated with three copy number states and CN jump sizes >0 indicated chromothripsis on two copies of the der(15;21) chromosome, consistent with chromothripsis on two sister chromatids of the der(15;21).

In summary, the der(15;21) chromosome of patient PD10009a likely underwent chromothripsis on two sister chromatids as the initiating rearrangement event. The derivative chromosome was later duplicated once or twice in association with circularization.



Supplementary Figure 23: Rearrangement metrics of chromosomes 15 and 21 for patient PD10009a. Copy number state distribution (A), copy number step distribution (B), copy number jump distribution (C). D shows a partial G-banded karyotype of the chromosomes 15 and 21 of this patient: from left to right are the normal chromosome 15, the normal chromosome 21 and the der(15;21) in the form of a large ring chromosome. In E, the left hand cell is simultaneously hybridized with whole chromosome paints specific for chromosomes 15 (WCP15, green) and 21 (WCP21, red). The green chromosome on the right and the red chromosome below indicate the normal chromosomes 15 and 21, respectively. To the left of the image the chromosomes showing interspersed patterns of red and green, manifesting as

yellow because they are closely apposed, indicates the der(15;21) with intermingling of chromosomes 15 and 21 material along its lengths. In the right hand image, the cell is hybridized with the TEL/AML1 ES probe (Abbott Molecular) showing multiple signals for *RUNX1* (red) clustered on a single chromosome. The discrete pair of red signals indicates *RUNX1* on the normal chromosome 21. The two pairs of green signals indicate the location of *ETV6* on the normal chromosomes 12. In A, the frequency distribution is shown as a histogram and the size of each segment is shown as small circles. Small horizontal lines depict median segment size of given copy number. In C, D – deletion type rearrangement link, TD – tandem duplication type, TT – tail to tail rearrangement, HH – head to head rearrangement.

5. Rearrangement validations

~95% of the called rearrangement were sent for validation through PCR and ~80% of them were validated as present in the tumour. This is consistent with the known failure rate of PCR for genomic rearrangements.

Supplementary methods

Sequence analysis

An unpublished somatic rearrangement analysis pipeline 'Brass' was used to align the sequencing data to NCBI37 reference genome (using BWA[31]), process raw alignments and call raw rearrangements. Raw rearrangements were filtered through several filters listed below to obtain the final rearrangement set. Several filters rely on removing false positive calls using normal panels. In addition, we developed two unpublished filtering strategies, *mapping difference (MD) score* and *insertion artefact (IA) score*, that are designed to capture the characteristics of false positive rearrangement calls caused by erroneous read mapping.

Brass outputs ranges of genomic positions for the two rearrangement breakpoints. A read pair is said to overlap with a raw rearrangement call, if the two reads each overlap one (but not the same) of the two ends of the rearrangement.

We have observed a high incidence of sequencing artefacts manifesting as fold-back type read pairs, and therefore used slightly different parameters for filtering fold-back and non-fold-back read pairs, as detailed below.

Filters

1. Raw rearrangements were cross-referenced against the sequencing data of two in-house panels of 21 and 24 normal DNA samples. The following procedure was carried out for both normal panels. Breakpoint ranges of the raw rearrangements were extended by 300 base pairs to account for insert size differences between the cancer sample and normal samples. Any raw rearrangements overlapping with more than two read pairs in either of the two normal panels were removed.

2. Rearrangements were cross-referenced against the sequencing data in the 1000 Genomes Project[32] (1000GP) as above, and a rearrangement call was removed if the rearrangement was
    a. fold-back type and overlapped with >6 1000GP read pairs.
    b. not fold-back type and overlapped with >1 1000GP read pairs.
3. MD score based filtering. BWA aligns every read to the genomic position with lowest number of mismatches. This can sometimes cause problems at repetitive and polymorphic loci. For example, let us assume that a read pair originates from an ancient L1 insertion, which we label *X*. The same read will map almost perfectly to many other instances of L1 in the reference genome due to the repetitiveness of L1. Let us further assume that in our cancer sample, there is a SNP at *X*, causing it to resemble most closely *another* instance of L1, *Y*, in the human reference genome. Now some sequencing reads that originated from *X* in our cancer sample will be mapped to *Y* at the reference genome, causing a read pairing pattern suggestive of a rearrangement between genomic positions of *X* and *Y*. To counteract this type of rearrangement calling artefact, we devised a strategy to measure the difference in DNA sequence between where a read was mapped and where it is *expected to map based on its mate's mapping position*. The individual steps of this method are described below.
    a. For each rearrangement call, all supporting read pairs are extracted from the BAM file of the sample.
    b. For each read pair, one read is used as an anchor while the MD score is calculated for the second read as detailed below.
    c. The unanchored read above is subjected to pairwise alignment against its mapping position (Perl package Algorithm::NeedlemanWunsch) allowing overhangs for the reference genome (+1 for match, -1 for mismatch, -3 for gap open and -1 for gap extend). Alignment is repeated at the genomic position where the unanchored read would be expected have originated from based on its anchored mate. The MD score of the read is then defined as [pairwise_alignment_score(mapping position) – pairwise_alignment_score(expected position based on mate)].
    d. The second read is set as anchor and MD score is calculated for the first read.
    e. MD score for the read pair is defined as min(MD score of read 1, MD score of read 2).
    f. Finally, MD score for the entire rearrangement is defined as $\min_i$(MD score of read pair i), where read pairs $i, i = 1,2, \ldots$ are the read pairs supporting the rearrangement.
    g. Patients PD7171a, PD10008a and PD7170a had read lengths of 50 and rearrangement calls of these samples were removed if their MD score was <35. Patients PD4117a, PD9020a, PD9021a, PD9022a, PD9023a had read lengths of up to 75bp and had MD score <50 as the criterion for removal. For patient PD10009a, the read pair level MDs were normalized by dividing the MDs to respective read lengths. Rearrangements calls were then removed if normalized MD < 0.7.

4. Insertion artefact based filtering. Some transposons are known to be active in the human genome. IA is designed to counteract rearrangement call artefacts arising from polymorphic insertions as retrotranspositions that are present in the sequenced sample but absent from the human reference genome. Let us assume that there is a germline retrotransposon insertion at position *X* in the sequenced sample that is absent from the reference genome. Now any reads originating from *X* are forcefully and erroneously mapped to another instance of this retrotransposon elsewhere in the genome. There are three signatures for such artefacts. 1) At one end of the rearrangement call, all reads will be mapped to a repetitive element, 2) at the other end of the rearrangement call, there will be other reads whose mates are mapped to the same repetitive element family elsewhere in the genome, and 3) for some reason the reads described in 1) tend to be clustered very closely together in their alignment position. Following steps were used to summarize these artefact signatures.
    a. The Repeatmasker data was downloaded from UCSC genome browser[33]. Repeat type to repeat family information was obtained from Genetic information research institute (www.girinst.org). The repetitive elements in Repeatmasker data were translated into their respective repetitive element families. In addition, all simple repeats (such as (TG)n) were labelled as "simple_repeat".
    b. The insert size distribution was estimated for every BAM file separately.
    c. For each rearrangement call, one end is anchored and the second is used to search for repetitive element signatures. Unless all the reads at the second end overlap with a same repeat masker element (30% overlap over the length of each read), IA score for this end was set to 0.
    d. A χ^2 test is conducted for the standard deviation of the 5'-end mapping positions of the reads on the unanchored end, using the estimated insert size distribution (b) as the null model. P-value from the test is recorded (*SD* score).
    e. Going back to the anchored end of the rearrangement, reads whose mates could have been derived from the putative inserted element (other than those supporting the putative rearrangement in question) are counted. A read was counted if (i) the putative polymorphic repetitive element insertion site was at the 3'-end of the read, (ii) mate of the read map to the same repetitive element family as the reads in (c) in the orientation that is consistent with the mapping orientation of the reads in (c). The total count is recorded (*IA_N*).
    f. SD score from (d) and IA_N from (e) are combined into a logistic model using the following parameters that were trained using independent training data (unpublished).

Intercept:	0.03328
SD:	-18.04808
IA_N:	0.95917

This procedure generates a final IA score for the unanchored rearrangement end in form of a probability that the rearrangement end is a polymorphic insertion related artefact.

g. The same procedure is repeated and the IA score is calculated for the second end of the rearrangement call. The final IA score for the rearrangement call is defined as max(IA score of end 1, IA score of end 2). Rearrangements with a final IA score of >0.75 were removed.

5. Fold-back type rearrangement calls (both ends mapping to the same genomic strand within 1kb) were removed if they were supported by fewer than four read pairs. Patient PD10009a had an extremely large number of fold-back rearrangements, so in this sample all fold-back rearrangements (two ends at same strand within 5kb) were removed regardless of the number of supporting read pairs.

Rearrangement validations

Primers mapping on either end of the reported structural variant in the appropriate orientation were designed and used by conventional PCR amplification. PCR reactions were performed in duplicate and amplicons were separated by agarose gel electrophoresis.

Copy number analysis

Read depths were calculated on variable size genomic windows[34]. Log_2 ratios were taken at read counts in the cancer sample versus read counts in a normal panel of 9 normal samples. GC-correction was performed by calculating the GC-content of each window, binning the GC windows into 1,000 equal-size bins and shifting the median $\log_2$ ratio of each bin to 0. GC-corrected data were median smoothed per chromosome using the R[35] function "runmed" with k=5. Smoothed windows were segmented using fastPCF[36]. Due to the high amount of noise in patient PD10009a, segment breakpoints for this sample were obtained from Affymetrix SNP6.0 DNA microarray data instead since the SNP6.0 analysis includes allelic information in segmenting (see below for analysis details). However, the patient PD10009a copy numbers of each segment were calculated using median of read depths from NGS data for consistency with other NGS samples.

To obtain correct estimates of copy number, we performed *in-silico* normal cell contamination removal. Given a tumour copy number segment of copy number cn_c and tumour cell fraction f, the observed $\log_2$ ratio $\log_2 R_{obs}$ is expected to be

$$\text{Log}_2 R_{obs} = \log_2 \left[\frac{f \times cn_c + (1-f) \times 2}{2}\right] = \log_2[(f \times cn_c + 2 - 2f)/2]. \quad (1)$$

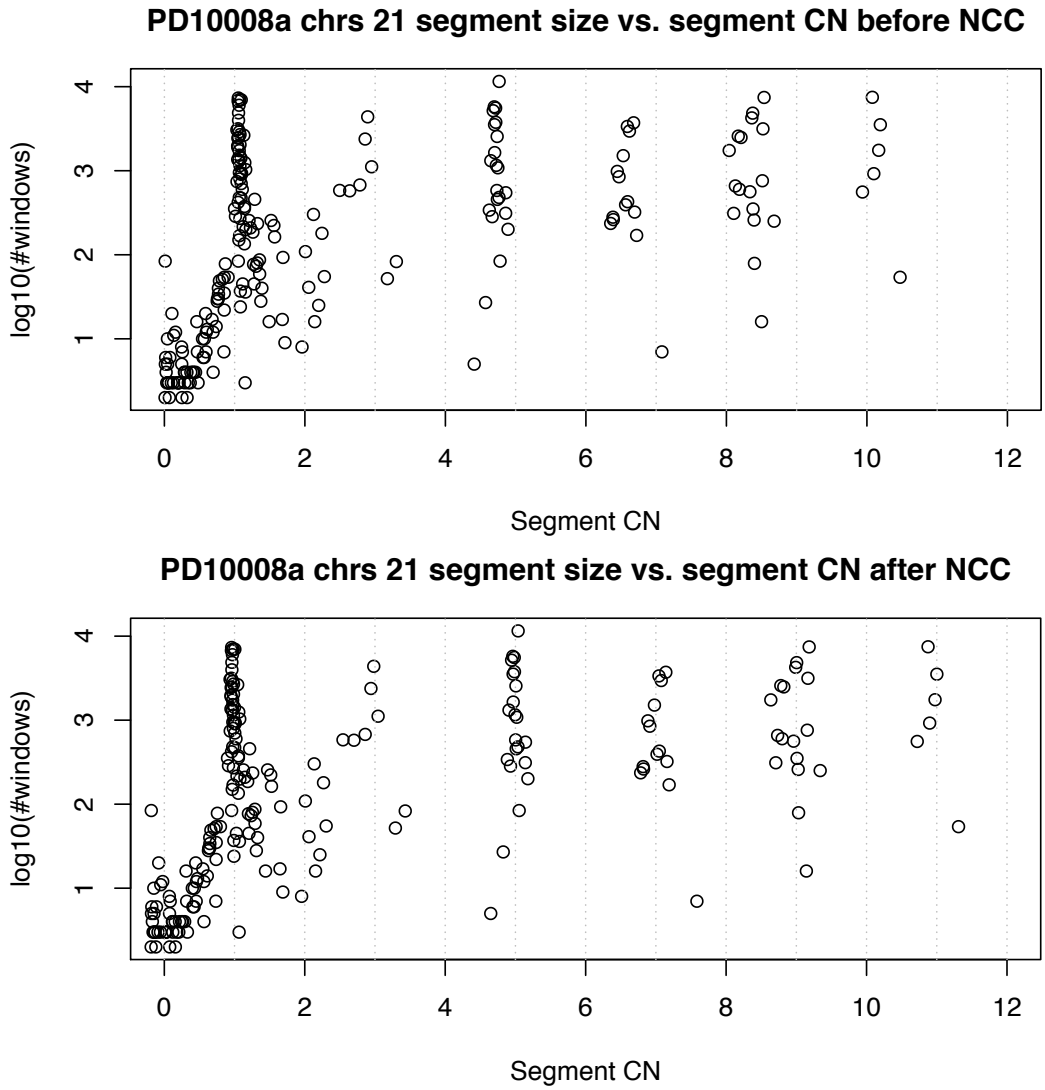
We can thus solve the estimated tumour CN

$$cn_c = 2 + 2 \times (r-1)/f. \quad (2)$$

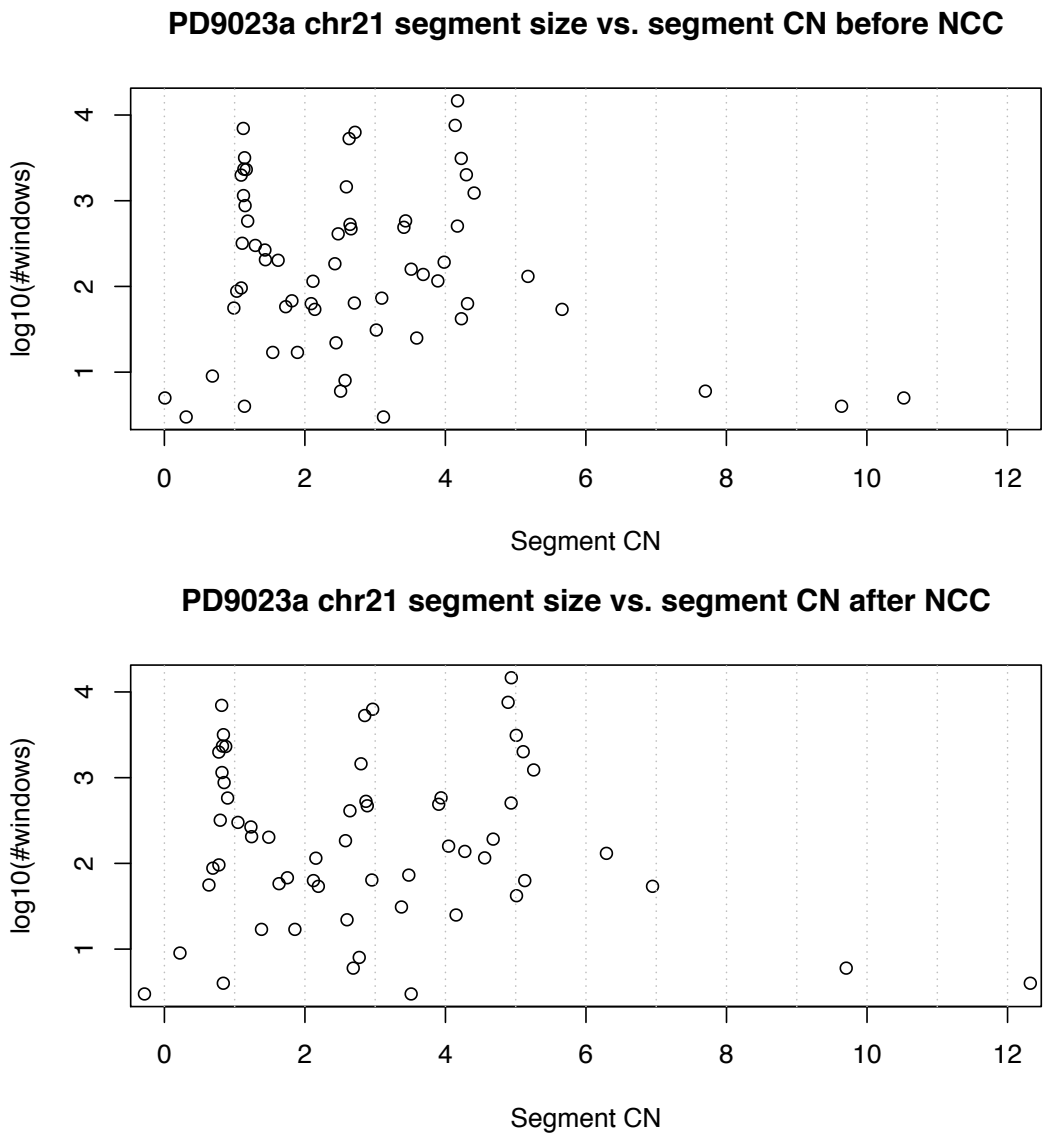
For each sample, we then tried different tumour fraction values f ranging from 0.5 to 1 at intervals of 0.01. For each value f, we calculated

$$\sum_i \left[cn_{c,i} - \text{round}\left(cn_{c,i}\right)\right]^2 \times \#\text{windows}_i, \quad (3)$$

where $cn_{c,i}$ is the adjusted copy number based on tumour cell fraction f and equation 2, #windows$_i$ is the number of windows in segment i and i goes over all segments in chromosome 21. The tumour fraction value f producing the smallest sum of squares in equation 3 is chosen as true tumour fraction and used to adjust the copy numbers. Examples of the effect of normal cell contamination correction are shown in Supplementary Figure 24 and Supplementary Figure 25.



Supplementary Figure 24: The effect of normal cell contamination removal to called copy numbers in patient PD10008a. Top and bottom panels show the copy number calls on chromosome 21 of patient PD10008a before and after normal cell contamination removal. Every data point represents one segment. X-axis shows the copy number and y-axis the 10-base logarithm of number of windows in a segment.



Supplementary Figure 25: The effect of normal cell contamination removal to called copy numbers in patient PD9023a. Top and bottom panels show the copy number calls on chromosome 21 of patient PD9023a before and after normal cell contamination removal. Every data point represents one segment. X-axis shows the copy number and y-axis the 10-base logarithm of number of windows in a segment.

Criteria for inferring chromothripsis

Breakpoint clustering

For this analysis, we removed any rearrangements that mapped within 1Mb of a centromere as defined by the UCSC genome browser 'hg19_gaps' table (http://hgdownload.soe.ucsc.edu/goldenPath/hg19/database/gap.txt.gz).

Statistical analysis for non-random breakpoint distribution was done using the exponential distribution for the null hypothesis as outlined in Ref. [30].

Randomness of rearrangement join orientations

The rearrangements were filtered as above. After that we used the R package 'EMT' to perform multinomial test to evaluate the statistical fit that all possible rearrangement orientations (tail-to-head, head-to-tail, tail-to-tail, head-to-head) occurring with identical chance of 0.25 in iAMP21, rob(15;21) chromosomes or in the rest of the genomes.

Chromothripsis effect

Per sample chromothripsis effect was computed from the segmented copy number data. Based on the sequences of rearrangements inferred in supplementary results section F, copy number alterations induced by BFB or large-scale duplication related amplifications were removed as described below. The coordinates were chosen according to copy number segmentation breakpoints.

Patient PD9023a

Chromothripsis effect not computed as the patient did not have chromothripsis.

Patient PD9022a

1. On segments with chromosome 21 position ≤ 23557176, set chromothripsis effect as unknown.
2. On segments between chromosome 21 coordinates 23557176 and 44089589, subtract copy number by one to remove copy number contribution from the wild-type chromosome 21.
3. Further, on segments between chromosome 21 coordinates 23557176 and 44089589:
    i. On copy number segments with copy number >6, set chromothripsis effect to 0
    ii. On copy number segments with copy numbers 2, 4 or 6, divide copy number by 2, then subtract 3 to set baseline copy number to 3 when chromothripsis happened.
    iii. On copy number segments with copy numbers 3 or 5, subtract copy number by 1, then divide copy number by 2, then subtract copy number further by 2 to set baseline copy number to 2 when chromothripsis happened.
4. On segments with chromosome 21 position > 44089589, set chromothripsis effect as unknown.

Patient PD4117a

1. On segments with chromosome 21 position ≤ 33931348, subtract copy number by 1 to remove copy number contribution from wild-type chromosome 21, then divide copy number by 2 to account for the late duplication event, then subtract by 2 to set baseline copy number to 2 when chromothripsis happened.
2. On segments with chromosome 21 position between 33931348 and 45923710, subtract copy number by 1 to remove copy number contribution from wild-type chromosome 21, then divide copy number by 2 to account for

the late duplication event, then subtract by 4 to set baseline copy number to 4 when chromothripsis happened.
3. On segments with chromosome 21 position > 45923710, set chromothripsis effect to unknown.

Patient PD9020a

1. On segments with chromosome 21 position ≤ 19019583, subtract copy number by 1 to remove copy number contribution from wild-type chromosome 21, then subtract by 2 to set baseline copy number to 2 when chromothripsis happened.
2. On segments with chromosome 21 position between 19019583 and 48042557, subtract copy number by 1 to remove copy number contribution from wild-type chromosome 21, then subtract by 4 to set baseline copy number to 4 when chromothripsis happened.
3. On segments with chromosome 21 position > 48042557, set chromothripsis effect to unknown.

Patient PD9021a

1. On segments with chromosome 21 position ≤ 17137902, subtract copy number by 1 to remove copy number contribution from wild-type chromosome 21, then subtract by 2 to set baseline copy number to 2 when chromothripsis happened.
2. On segments with chromosome 21 position between 17137902 and 24101175, subtract copy number by 1 to remove copy number contribution from wild-type chromosome 21, then subtract by 4 to set baseline copy number to 4 when chromothripsis happened.
3. On segments with chromosome 21 position between 24101175 and 44314457, subtract copy number by 1 to remove copy number contribution from wild-type chromosome 21, then divide copy number by 2 to account for the late duplication event, then subtract by 4 to set baseline copy number to 4 when chromothripsis happened.
4. On segments with chromosome 21 position > 44314457, set chromothripsis effect to unknown.

Patient PD7171a

We assume that in patient PD7171a the initiating rearrangement event was chromothripsis on two sister chromatids. However, we do not know whether copy number states 3 and 4 of the der(15;21) chromosome were acquired through several duplication events or through a whole-chromosome duplication (such as isochromosome formation) followed by deletions.

In both scenarios, copy numbers (excluding the wild-type chromosome 15 and 21) 3 and 4 are parsimoniously generated from segments of copy number 2 after chromothripsis. In the former scenario, copy number 2 is parsimoniously generated from regions of copy number 2 after chromothripsis while in the latter scenario they are parsimoniously from regions of copy number 1 after chromothripsis. We therefore extracted the chromothripsis effects as below.

1. Set segments of (total) copy number 1 as chromothripsis effect -2.
2. Set segments of copy number 2 as chromothripsis effect -1.
3. Set segments of copy number 3 as chromothripsis effect -1.5.
4. Set segments of copy number > 4 as chromothripsis effect 0.

Patient PD1008a

For all segments, subtract copy number by 1 to remove copy number contribution of the wild-type chromosomes 15 and 21, then divide by 4 to remove the effect of whole-chromosome duplications, then subtract by 2 to set baseline copy number to 2 when chromothripsis occurred.

Patient PD7170a

For all segments, subtract copy number by 1 to remove copy number contribution of the wild-type chromosomes 15 and 21, then divide by 2 to remove the effect of the whole-chromosome duplication, then subtract by 2 to set baseline copy number to 2 when chromothripsis occurred.

Patient PD10009a

For all segments, subtract copy number by 1 to remove copy number contribution of the wild-type chromosome 15 and 21, then divide by 3 to remove the effect of whole-chromosome duplication, then subtract by 2 to set baseline copy number to 2 when chromothripsis occurred.

Summarizing chromothripsis effect

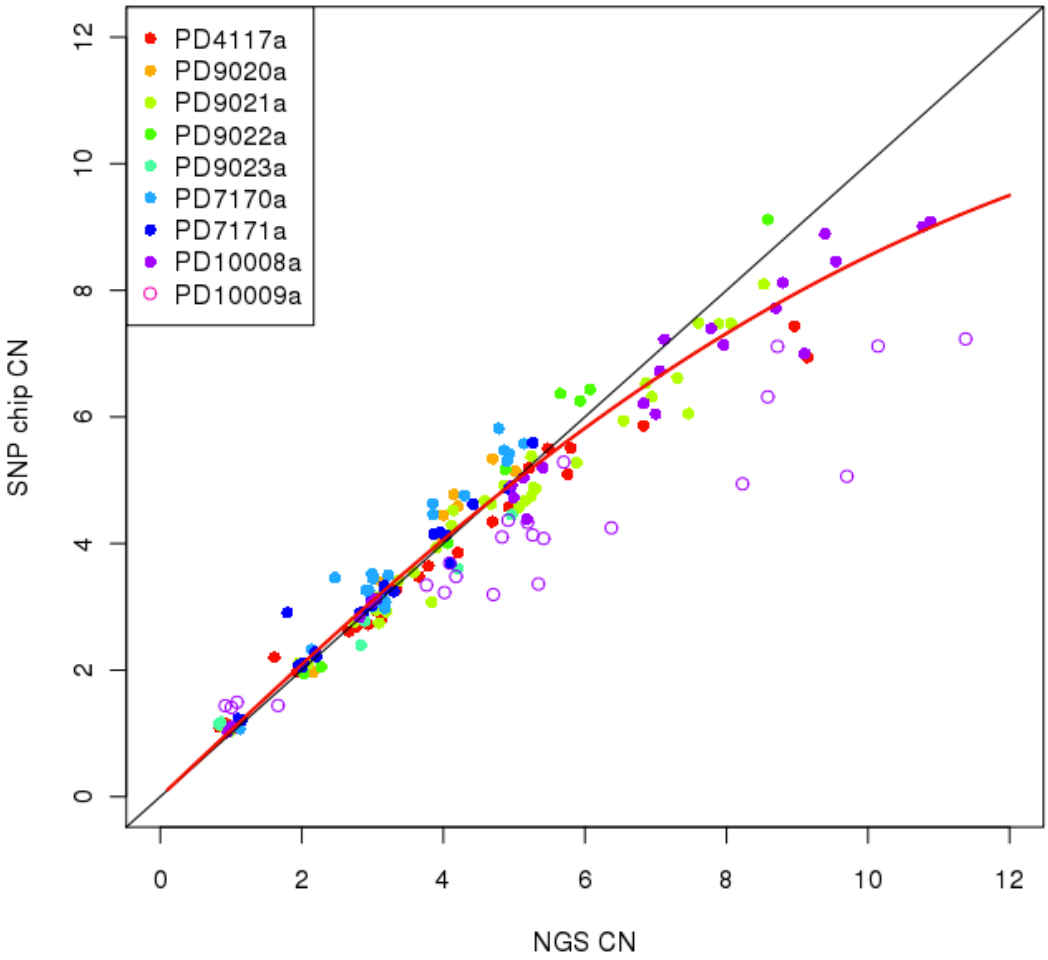
Sample-level copy number segments were converted into chromothripsis effect segments as explained above. Chromosome 21 was divided into non-overlapping 100kb windows and average chromothripsis effect was computed over every window and for every sample.

SNP6.0 chip analysis

SNP6.0 data were produced using Affymetrics arrays (Aros, Denmark) and analyzed using Genotyping Console software (Affymetrix). Segmentation was done using ASCAT[37]. We used the samples with both sequencing and SNP6.0 data available (excluding patient PD10009a as the DNA quality was poor) to derive a saturation model for SNP6.0 signal saturation at high intensities (Supplementary Figure 26). We observed that SNP6.0 signal follows the sequencing depth based copy number estimate linearly up to copy number 6, after which it starts to saturate. We fit a logistic function

$$CN_{SNP} = asym/[1 + \exp(CN_{NGS} \times scal)] - asym/2$$

and obtained $asym = -24.52$ and $scal = 5.81$. The fitted model was used to correct for copy number estimates obtained from ASCAT.



Supplementary Figure 26: A signal saturation model for SNP6.0 data. Copy number derived from SNP6.0 data increases linearly with actual copy number until ~6 copies where it starts to saturate. The estimated saturation curve is plotted in read. Data from patient PD10009a (open circles) were not used in model construction.

Expression analysis

The expression data has been published previously[38]. The dataset consists of 8 iAMP21 and 81 non-iAMP21 ALL samples. Probe sets flagged as 'Absent' in ≥45 of the samples were excluded from analysis. The coefficients of variation were computed for the remaining probe sets among the iAMP21 and non-iAMP21 samples, resulting in two coefficients of variation per probe set. For each probe set, the mean of the two coefficients of variation was then used as its weight to calculate the weighted average expression ratio between iAMP21 and non-iAMP21 groups over all probe sets of each gene.

Simulations

We developed a chromosomal rearrangement simulation framework study the characteristics of the chromosomes undergoing different sequences of rearrangement events. Chromosomes of size 1Mb were used as 'wild-type'

chromosomes and two rearrangement breakpoints are never allowed to fall within 1kb from each other. Source code for the simulation framework is available upon request. The functional definitions of the rearrangement events used in this simulation framework are detailed below.

Mutation functions

Deletion

The function takes a maximum deletion length as an input argument and generates two random deletion breakpoints. The breakpoints are generated from a truncated normal distribution that is centred at (max deletion length)/2 and scaled such that -2 standard deviations (SD) corresponds to 0 and +2 SDs corresponds to (max deletion length). The region between two to breakpoints is deleted in the output chromosome.

Tandem duplication

The function takes a maximum deletion length as an input argument and generates two random duplication breakpoints. The breakpoints are generated from a truncated normal distribution that is centred at (max deletion length)/2 scaled such that -2 SDs corresponds to 0 and +2 SDs corresponds to (max deletion length). The region between two to breakpoints is duplicated in tandem in the output chromosome.

Telomeric double stranded breakage

A breakpoint is generated randomly between ¾ of the length of the input chromosome and the q-terminus. The region between the generated breakpoint and the q-terminus of the input chromosome is deleted in the output chromosome. The breakpoints are generated from a truncated normal distribution scaled such that -2 SDs corresponds to ¾ input chromosome length and +2 SDs corresponds to the q-terminus.

Fusion of chromosome ends

The input chromosome is duplicated and fused to itself in tail-to-tail orientation.

Double-stranded break at the center of the chromosome

A breakpoint is randomly generated between ¼ and ¾ of the input chromosome length, and the region between the breakpoint and the q-terminus of the chromosome is deleted in the output chromosome. The breakpoints are generated from a truncated normal distribution scaled such that -2 SDs corresponds to ¼ chromosome length and +2 SDs corresponds to ¾ chromosome length.

Chromothripsis

First, *k* breakpoints are randomly generated from a uniform distribution, where *k* is generated from Poisson distribution with an expected value of 20. Input chromosome is broken into segments at the breakpoints, and each segment is lost with a probability of 0.4. Finally the segments are joined back together in random order and orientation.

Simultaneous chromothripsis at two sister chromatids

Same as chromothripsis simulation, but now the input chromosome is first duplicated, and both duplicated copies are shattered and joined together simultaneously.

Special cases in simulating rearrangement sequences

- Double-stranded DNA breakage (DSB) was always repeated until it produced a chromosome that was longer than half of the input chromosome size. This requirement is enforced to ensure that every simulated fusion-breakage cycle results in increase of chromosomal material.
- A process of two BFB cycles is simulated as a telomeric DSB, followed by fusion of chromosome ends, followed by DSB at the centre of the derivative chromosome, and finally completed by another fusion event.
- When simulating chromothripsis followed by two tandem duplications, the maximum tandem duplication size was set to ½ input chromosome length.

Predicting chromothripsis effect from deletions in cancer

Three large cancer deletion datasets were compiled from Kim et al. 2013 (http://compbio.med.harvard.edu/metacgh/segmentations/segmentation8227.CBS.hg19.seg)[39], Beroukhim et al. 2010 (ftp://ftp.broadinstitute.org/pub/cancer/gcnmp/Summary_data/Segmented_data/tumorscape_100217.seg)[40] and an unpublished dataset.

Chromosome 21 17-46Mb was divided into 1Mb windows. In Kim *et al.* and Beroukhim *et al.* datasets, regions with log ratio < log2(1.6/2) were treated as deletions, whereas in the unpublished dataset regions of deletions were given by the analysis. For each of the three cancer copy number datasets, the average fraction of each window deleted over all samples was computed.

The average chromothripsis effect was computed for the same 1Mb windows. We used Wald test to test whether the average deletion fraction of a 1Mb window had predictive power for chromothripsis effect of the same window. In the null model, chromothripsis of a window at i Mb y_i is only dependent on autocorrelation.

$$y_i = \alpha y_{i-1} + \varepsilon_i,$$

where ε_i are independent, identically distributed random variables with a common variance. In the alternative model, y_i is also linearly dependent on average fraction of deletions in one of the three cancer copy number datasets.

$$y_i = \alpha y_{i-1} + \beta x_i + \varepsilon_i,$$

where x_i is average deletion fraction at i Mb in one of the three cancer datasets. P-values were computed independently for all three cancer copy number datasets.

To control for general chromosome arm-level copy number biases, nine other chromosome arms with similar size to 21q were selected (Supplementary Table 5). Similar to 21q, average deletion fractions over a 29 Mb region centered at the each chromosome arm was computed for each cancer copy number dataset. For all the p-arms, the per-window average deletion fraction series were reversed so as to have

the windows proceeding from more centromeric to more telomeric, as in the 21q chromothripsis data.

Statistical significance of predicting the chromothripsis effect values were computed for each of these "mock" average deletion fraction series against for each cancer copy number dataset.

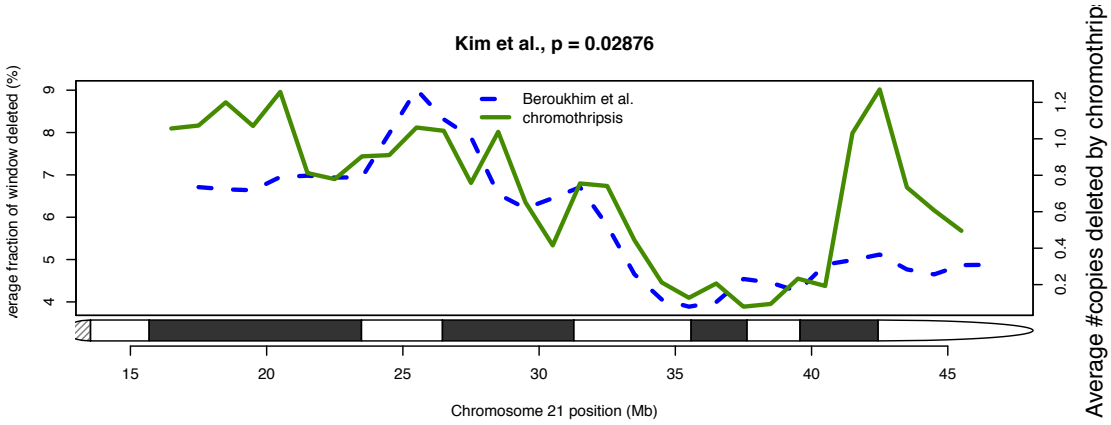
Supplementary Table 5: Chromosome arms whose average deletion fractions were used to predict chromothripsis effect on 21q as control experiments.

Chromosome arm	Chromosome arm length (Mb)
5p	48.8
8p	45.6
9p	49.0
10p	40.0
12p	35.8
16p	36.6
19q	32.6
20q	35.5
22q	36.6

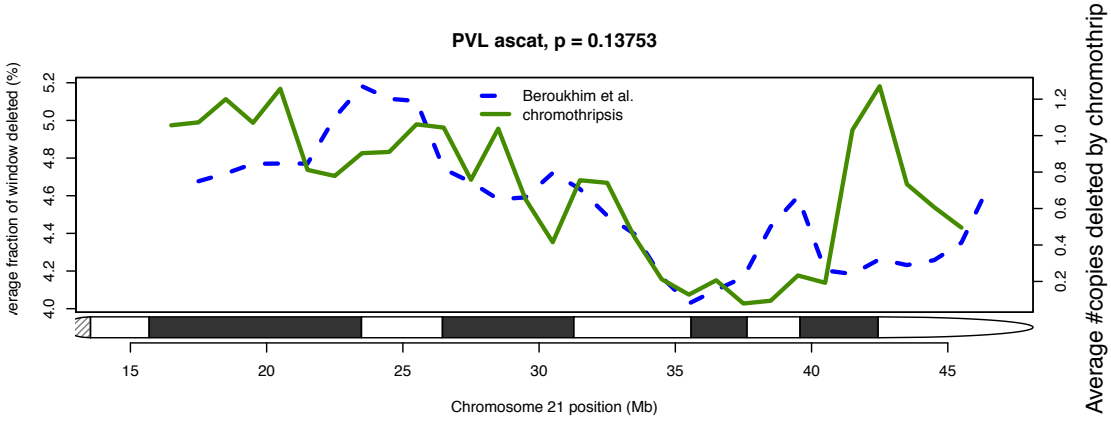
Plots

Chromosomal ideograms were generated using the R Bioconductor package quantsmooth[41]. Diagrams were generated using R and Integrative genomics viewer[42,43].

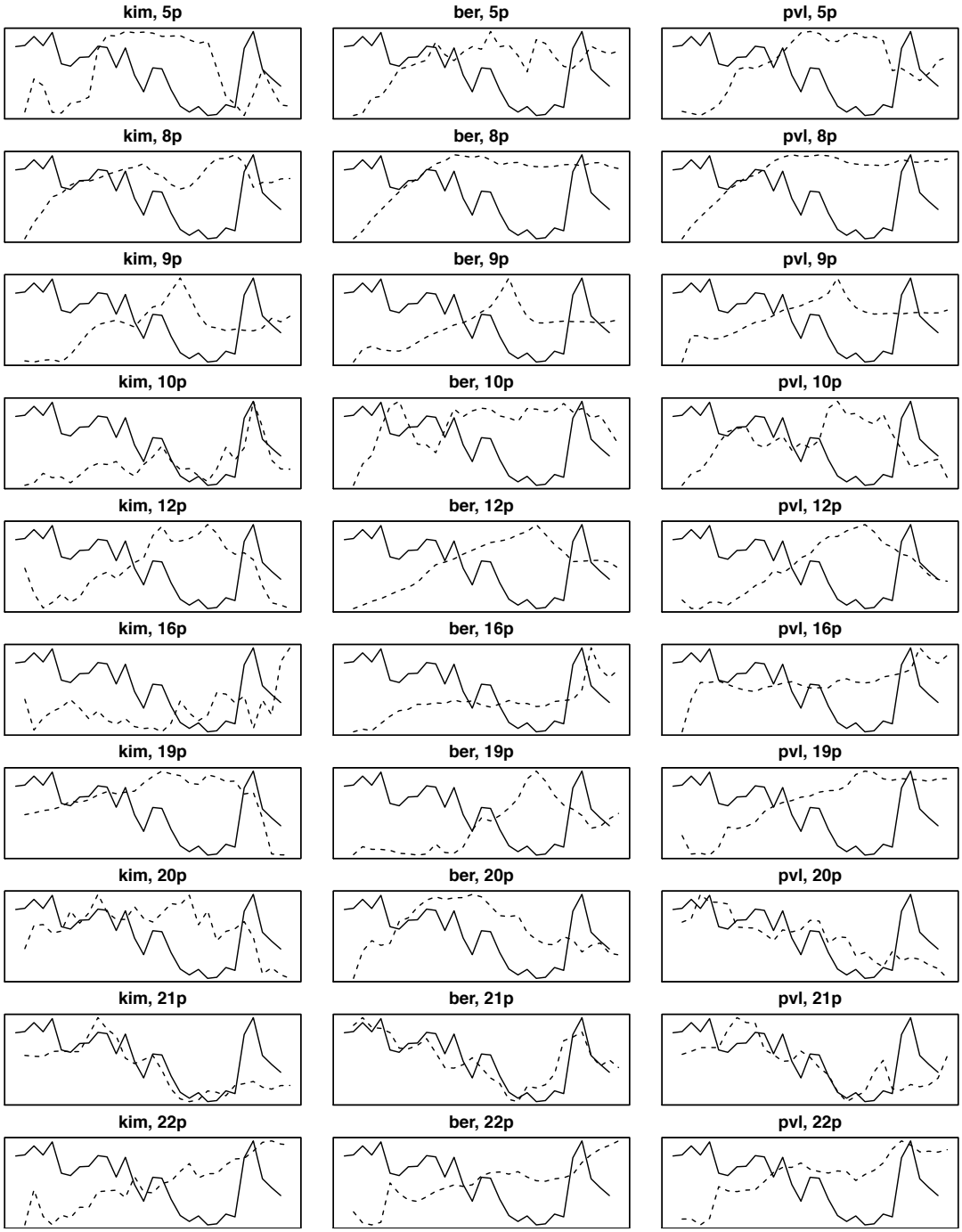
Supplementary Figures 27-30



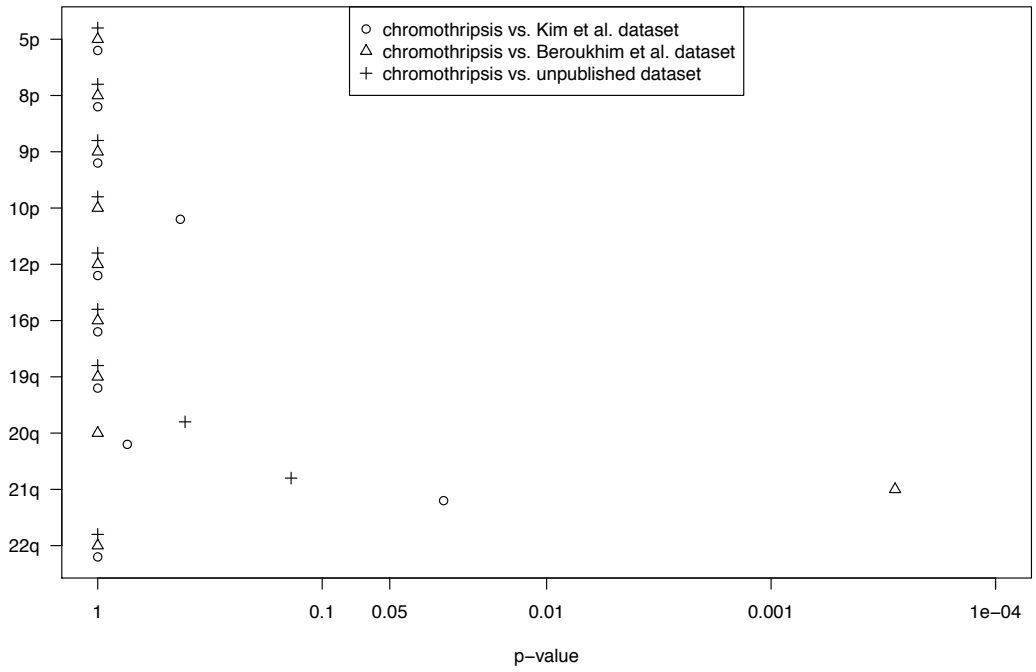
Supplementary Figure 27: Chromothripsis effect on chromosome 21 vs. average deletion fraction series from Kim et *al*. dataset.



Supplementary Figure 28: Chromothripsis effect on chromosome 21 vs. average deletion fraction series from an unpublished dataset.



Supplementary Figure 29: Chromothripsis effect only mirrors the deletion landscape of chromosome 21. Chromothripsis effect of chromosome 21 (solid line) vs. average deletion fraction series on ten chromosome arms (Supplementary Table 5) in each of the three cancer copy number datasets. All but 21q were used as control experiments to the observed correlation between average deletion series and chromothripsis effect on 21q. Corresponding *p*-values are shown in Supplementary Figure 30.



Supplementary Figure 30: *P*-values of correlations between chromosome arm average deletion fraction series and 21q chromothripsis effect. The corresponding graphs are shown in Supplementary Figure 29.

Supplementary table 6

Supplementary Table 6: Sequencing statistics.

Sample	Run	Lane	Read Length	# Read Pairs	Gbp Sequenced
PD4117a	4705	5	37	36697201	2.715592874
PD4117a	4705	6	37	37123727	2.747155798
PD4117a	4705	7	37	31247328	2.312302272
PD4117a	4705	8	37	34029969	2.518217706
PD4117a	8959	1	75	117593340	17.639001
PD9020a	7928	5#146	50	21357626	2.1357626
PD9020a	8899	1#146	75	28821497	4.32322455
PD9020a	8899	2#146	75	28786171	4.31792565
PD9021a	7928	5#147	50	48892219	4.8892219
PD9021a	8899	1#147	75	57558533	8.63377995
PD9021a	8899	2#147	75	57468666	8.6202999
PD9022a	7928	5#148	50	56142110	5.614211
PD9022a	8899	1#148	75	61345075	9.20176125
PD9022a	8899	2#148	75	61257680	9.188652
PD9023a	7928	5#149	50	47402101	4.7402101
PD9023a	8899	1#149	75	53376627	8.00649405
PD9023a	8899	2#149	75	53303413	7.99551195
PD7170a	6471	6	50	66024547	6.6024547
PD7170a	6501	1	50	105757485	10.5757485
PD7170a	6501	2	50	88130527	8.8130527
PD7171a	6471	7	50	108233705	10.8233705
PD10008a	8994	7	50	166052909	16.6052909
PD10009a	8947	4	75	131908366	19.7862549
PD10009a	8994	8	50	129958134	12.9958134

References

1 Harewood, L. *et al.* Amplification of AML1 on a duplicated chromosome 21 in acute lymphoblastic leukemia: a study of 20 cases. *Leukemia* **17**, 547-553 (2003).

2 Soulier, J. *et al.* Amplification of band q22 of chromosome 21, including AML1, in older children with acute lymphoblastic leukemia: an emerging molecular cytogenetic subgroup. *Leukemia* **17**, 1679-1682 (2003).

3 Alvarez, Y., Coll, M. D., Bastida, P., Ortega, J. J. & Caballin, M. R. AML1 amplification in a child with acute lymphoblastic leukemia. *Cancer Genetics and Cytogenetics* **140**, 58-61 (2003).

4 Dube, I. D. & el Solh, H. An apparent tandem quadruplication of chromosome 21 in a case of childhood acute lymphoblastic leukemia. *Cancer Genetics and Cytogenetics* **23**, 253-256 (1986).

5 Le Coniat, M., Romana, S. P. & Berger, R. Partial chromosome 21 amplification in a child with acute lymphoblastic leukemia. *Genes.Chromosomes.Cancer.* **14**, 204-209 (1995).

6 Busson-Le Coniat, M., Nguyen, K. F., Daniel, M. T., Bernard, O. A. & Berger, R. Chromosome 21 abnormalities with AML1 amplification in acute lymphoblastic leukemia. *Genes Chromosomes.Cancer* **32**, 244-249 (2001).

7 Baialardo, E. M., Felice, M. S., Rossi, J., Barreiro, C. & Gallego, M. S. Tandem triplication and quadruplication of chromosome 21 in childhood acute lymphoblastic leukemia. *Cancer Genetics and Cytogenetics* **92**, 43-45 (1996).

8 Niini, T., Kanerva, J., Vettenranta, K., Saarinen-Pihkala, U. M. & Knuutila, S. AML1 gene amplification: a novel finding in childhood acute lymphoblastic leukemia. *Haematologica* **85**, 362-366 (2000).

9 Dal Cin, P. *et al.* Amplification of AML1 in childhood acute lymphoblastic leukemias. *Genes Chromosomes.Cancer* **30**, 407-409 (2001).

10 Mathew, S., Rao, P. H., Dalton, J., Downing, J. R. & Raimondi, S. C. Multicolor spectral karyotyping identifies novel translocations in childhood acute lymphoblastic leukemia. *Leukemia* **15**, 468-472 (2001).

11 Morel, F. *et al.* AML1 amplification in a case of childhood acute lymphoblastic leukemia. *Cancer Genetics and Cytogenetics* **137**, 142-145 (2002).

12 Penther, D. *et al.* Amplification of AML1 gene is present in childhood acute lymphoblastic leukemia but not in adult, and is not associated with AML1 gene mutation. *Leukemia* **16**, 1131-1134 (2002).

13 Moorman, A. V. *et al.* Risk directed treatment intensification significantly reduces the risk of relapse among children with acute lymphoblastic leukaemia and intrachromosomal amplification of chromosome 21. . *Journal of Clinical Oncology* **in press** (2013).

14 Robinson, H. M. *et al.* Amplification of AML1 in acute lymphoblastic leukemia is associated with a poor outcome. *Leukemia* **17**, 2249-2250 (2003).

15 Moorman, A. V. *et al.* Prognosis of children with acute lymphoblastic leukemia (ALL) and intrachromosomal amplification of chromosome 21 (iAMP21). *Blood* **109**, 2327-2330 (2007).

16 Attarbaschi, A. *et al.* Minimal residual disease values discriminate between low and high relapse risk in children with B-cell precursor acute lymphoblastic leukemia and an intrachromosomal amplification of chromosome 21: the Austrian and German acute lymphoblastic leukemia Berlin-Frankfurt-Munster (ALL-BFM) trials. *J Clin Oncol* **26**, 3046-3050 (2008).

17 Heerema, N. A. *et al.* Intrachromosomal Amplification of Chromosome 21 Is Associated With Inferior Outcomes in Children With Acute Lymphoblastic Leukemia Treated in Contemporary Standard-Risk

Children's Oncology Group Studies: A Report From the Children's Oncology Group. *Journal of Clinical Oncology* **in press** (2013).

18 Rand, V. *et al.* Genomic characterization implicates iAMP21 as a likely primary genetic event in childhood B-cell precursor acute lymphoblastic leukemia. *Blood* **117**, 6848-6855 (2011).

19 Robinson, H. M., Harrison, C. J., Moorman, A. V., Chudoba, I. & Strefford, J. C. Intrachromosomal amplification of chromosome 21 (iAMP21) may arise from a breakage-fusion-bridge cycle. *Genes Chromosomes.Cancer* **46**, 318-326 (2007).

20 Strefford, J. C. *et al.* Complex genomic alterations and gene expression in acute lymphoblastic leukemia with intrachromosomal amplification of chromosome 21. *Proc.Natl.Acad.Sci.U.S.A* **103**, 8167-8172 (2006).

21 Kuchinskaya, E. *et al.* Tiling-resolution array-CGH reveals the pattern of DNA copy number alterations in acute lymphoblastic leukemia with 21q amplification: the result of telomere dysfunction and breakage/fusion/breakage cycles? *Leukemia* **21**, 1327-1330 (2007).

22 Sinclair, P. B. *et al.* Analysis of a breakpoint cluster reveals insight into the mechanism of intrachromosomal amplification in a lymphoid malignancy. *Hum Mol Genet* **20**, 2591-2602 (2011).

23 Harrison, C. J. Cytogenetics of paediatric and adolescent acute lymphoblastic leukaemia. *Br J Haematol* **144**, 147-156 (2009).

24 Harrison, C. J. *et al.* Detection of prognostically relevant genetic abnormalities in childhood B-cell precursor acute lymphoblastic leukaemia: recommendations from the Biology and Diagnosis Committee of the International Berlin-Frankfurt-Munster study group. *Br J Haematol* **151**, 132-142, doi:10.1111/j.1365-2141.2010.08314.x (2010).

25 Hamerton, J. L., Canning, N., Ray, M. & Smith, S. A cytogenetic survey of 14,069 newborn infants. I. Incidence of chromosome abnormalities. *Clin Genet* **8**, 223-243 (1975).

26 Jacobs, P. A., Browne, C., Gregson, N., Joyce, C. & White, H. Estimates of the frequency of chromosome abnormalities detectable in unselected newborns using moderate levels of banding. *J Med Genet* **29**, 103-108 (1992).

27 Stiller, C. A., Kroll, M. E., Boyle, P. J. & Feng, Z. Population mixing, socioeconomic status and incidence of childhood acute lymphoblastic leukaemia in England and Wales: analysis by census ward. *Br J Cancer* **98**, 1006-1011, doi:10.1038/sj.bjc.6604237 (2008).

28 Ensor, H. M. *et al.* Demographic, clinical, and outcome features of children with acute lymphoblastic leukemia and CRLF2 deregulation: results from the MRC ALL97 clinical trial. *Blood* **117**, 2129-2136, doi:10.1182/blood-2010-07-297135 (2011).

29 Harrison, C. J. *et al.* Intrachromosomal Amplification of Chromosome 21 (iAMP21): Cytogenetic characterization and outcome. An International

Study on behalf of the Ponte di Legno International Workshop in Childhood Acute Lymphoblastic Leukemia. *Leukemia* **Submitted** (2013).

30 Korbel, J. O. & Campbell, P. J. Criteria for inference of chromothripsis in cancer genomes. *Cell* **152**, 1226-1236, doi:10.1016/j.cell.2013.02.023 (2013).

31 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760, doi:10.1093/bioinformatics/btp324 (2009).

32 A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061-1073, doi:10.1038/nature09534 (2010).

33 Meyer, L. R. *et al.* The UCSC Genome Browser database: extensions and updates 2013. *Nucleic Acids Res* **41**, D64-69, doi:10.1093/nar/gks1048 (2013).

34 Campbell, P. J. *et al.* Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing. *Nat Genet* **40**, 722-729, doi:10.1038/ng.128 (2008).

35 R: A language and environment for statistical computing (R Foundation for Statistical Computing, Vienna, Austria, 2010).

36 Baumbusch, L. O. *et al.* Comparison of the Agilent, ROMA/NimbleGen and Illumina platforms for classification of copy number alterations in human breast tumors. *BMC Genomics* **9**, 379, doi:10.1186/1471-2164-9-379 (2008).

37 Van Loo, P. *et al.* Allele-specific copy number analysis of tumors. *Proc Natl Acad Sci U S A* **107**, 16910-16915, doi:10.1073/pnas.1009843107 (2010).

38 Strefford, J. C. *et al.* Complex genomic alterations and gene expression in acute lymphoblastic leukemia with intrachromosomal amplification of chromosome 21. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 8167-8172, doi:10.1073/pnas.0602360103 (2006).

39 Kim, T. M. *et al.* Functional genomic analysis of chromosomal aberrations in a compendium of 8000 cancer genomes. *Genome Res* **23**, 217-227, doi:10.1101/gr.140301.112 (2013).

40 Beroukhim, R. *et al.* The landscape of somatic copy-number alteration across human cancers. *Nature* **463**, 899-905, doi:10.1038/nature08822 (2010).

41 Eilers, P. H. & de Menezes, R. X. Quantile smoothing of array CGH data. *Bioinformatics* **21**, 1146-1153, doi:10.1093/bioinformatics/bti148 (2005).

42 Robinson, J. T. *et al.* Integrative genomics viewer. *Nat Biotechnol* **29**, 24-26, doi:10.1038/nbt.1754 (2011).

43 Thorvaldsdottir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform*, doi:10.1093/bib/bbs017 (2012).