Title: The genetic sequences prone to copy number variation and single nucleotide

2 polymorphism are linked to the repair of the poisoned DNA topoisomerase II

Authors:

Chuo Jiang^{1,2,¶,#}, Cong Ma^{3,¶}, Detao Wang^{4,¶}, Li Liu⁴, Chunxiu Zhang⁴, Fuxue Chen¹, Jiaxi
Wu^{2,5,6,*}

Affiliations:

¹School of Life Sciences, Shanghai University, 99 Shangda Road, Shanghai, China.

²Central Laboratories, Xuhui Central Hospital, Shanghai Clinical Research Center, Chinese
Academy of Sciences, 966 Middle Huaihai Road, Shanghai, China.

2 ³Department of Endocrinology, Xuhui Central Hospital, 966 Middle Huaihai Road, Shanghai,
China.

⁴Shanghai Biotechnology Corporation, 151 Libing Road, Zhangjiang Hi-Tech Park, Pudong,
Shanghai, China.

⁵Shanghai Institute of Nutrition and Health, Chinese Academy of Sciences, 320 Yueyang
Road, Shanghai, China.

⁶Department of Pathology and Laboratory Medicine, Brody School of Medicine, East
Carolina University, 600 Moye Boulevard, Greenville, North Carolina, USA.

2 [#]Current address: Shanghai Biotechnology Corporation, 151 Libing Road, Zhangjiang
2 Hi-Tech Park, Pudong, Shanghai, China.

22 [¶]These three authors contributed equally to this work.

2

2 *Corresponding author

2 Email: jxwu@scrc.ac.cn (JW)

Abstract

TOP2-poisoning bioflavonoids and pesticides are linked to the copy number variation-related autism and chromosome translocation-related leukemia. On the other hand, the poisoned DNA topoisomerase II (TOP2) can lead to chromosome aberration. However, except a limited number of genes such as the *MLL* fusion, other poisoned TOP2-targeted genes, as well as their relationships with any specific diseases, are not defined. We applied the γ H2A.X antibodies to genome-widely immunoprecipitate the chromatins that were associated with the repair of the TOP2 poison etoposide-induced DNA double strand breaks. We identified many transcribable protein- and nonprotein-coding DNA sequences that are the candidates of or associated with many gene copy number variation- and/or single nucleotide polymorphism-associated diseases, including but not limited to microdeletion and microduplication syndromes (which are phenotypically presented as developmental, autistic, neurological, psychiatric, diabetic, autoimmune, and neoplastic diseases among many others) as well as stature, obesity, metabolic syndrome, hypertension, coronary artery disease, ischemic stroke, aortic aneurysm and dissection, leukemia, cancer, osteoporosis, Alzheimer disease, Parkinson disease, and Huntington disease. Our data raise the possibility that the poisoned TOP2 might be linked to the specific genetic alterations contributing to these diseases, additional to the known copy number variation-related autism and chromosome translocation-related leukemia. According to our and others' data, we propose a model that may interpret the features, such as mosaicism, polygenic traits and pleiotropy, of these diseases.

Author Summary

For the past several decades, the morbidity rate of many diseases, including autism, mental disorders, cancer, cardiovascular diseases, diabetes, and senile dementia, has world-widely been rising. Analysis of the genome of the patients and their family members has identified the genes, whose alterations, so called copy number variation (CNV) and single nucleotide polymorphism (SNP), contribute to the diseases. Moreover, the CNVs and SNPs are *de novo*, that is, they have occurred only in the recent generations. Epidemiologically, this indicates that for the past several decades, there have existed some unknown world-wide etiologies to which human beings are exposed. If the etiologies are identified, avoiding human's exposure may reduce the morbidity of the diseases. We have found that the repair of the poisoned topoisomerase II involves many genes that contribute to the aforementioned diseases. As the topoisomerase II is known to be located at the genomic sites where the disease-associated CNVs occur, as the poisoned topoisomerase II is susceptible to chromosome aberration, and as the topoisomerase II poisons, such as dietary bioflavonoids, are widely distributed in the environment, our data raise the yet-to-be-confirmed possibility that the environmental topoisomerase II poisons might etiologically contribute to many CNV-associated diseases.

Introduction

γ H2A.X is a DNA repair-involved, post-translationally modified form of histone variant

H2A.X [1-4]. When a DNA double strand break (DSB) is elicited, H2A.X is phosphorylated on its serine 139 to generate γ H2A.X as an early cellular DNA repair response [1-4]. γ H2A.X molecules can spread up to megabases along the DNA flanking the DSB [1-4]. Previous studies have shown that the endogenous DSB causes the *in situ* formation of γ H2A.X from the H2A.X that is located around the sub-telomeric region and transcription starting site in the Jurkat cell [5]. On the other hand, the exogenous DSB induced by 10Gy ionizing irradiation leads to the distribution of γ H2A.X at the pausing site of RNA polymerase II within the gene body in the resting normal CD4 cell [5].

Human DNA topoisomerase II (TOP2) is an enzyme that relieves the topological strain generated during replication, transcription and other DNA transactions [6-8]. The enzyme may be poisoned by both distorted DNA structures, such as abasic DNA and DNA adducts, and many endogenous and exogenous agents, such as reactive oxygen species and bioflavonoids [6-8]. The poisoned TOP2 is converted by the cell into a TOP2-linked DSB known as cleavable complex [6-8], which subsequently triggers a series of cellular DNA repair responses, including the formation of γ H2A.X [3,9,10].

Additional to the DSB, the poisoned TOP2 may also cause a DNA single strand break (SSB) [11]. However, the SSB does not seem to be related to the formation of γ H2A.X and cytotoxicity [12]. More importantly, the poisoned TOP2-induced chromosome aberration, on which the current study focuses, is considered to result from the DSB, not the SSB [13].

TOP2 poisons may cause chromosome aberration [6-8,14,15]. Additionally, the TOP2-poisoning anticancer drug etoposide, bioflavonoids and pesticides have been linked to

the chromosome translocation-related leukemia and/or gene copy number variation (CNV)-related autism [7,13,16-18]. However, except a limited number of genes such as the leukemia-related *MLL* fusion, other mutated genes resulting from the TOP2 poison-induced DNA damage and, particularly its repair, are not defined. Nor are their relationships with any specific diseases.

Mechanistically, the formation of the *MLL* fusion is considered to be caused by the DSB induced by the poisoned TOP2 [13], apoptotic caspase-activated DNase (CAD) [19,20] and/or other mechanisms [21,22]. However, clinically, patients who have the *MLL* fusion and develop leukemia, often have a history of receiving the TOP2-poisoning anticancer drugs [23].

Additional to the autism, CNVs are attributed to a group of diseases named as microdeletion and microduplication syndromes (or genomic disorders) which are characterized by the deletion or duplication of a genomic segment that often encompasses a number of transcribable protein- or nonprotein-coding sequences [24-28]. (Hereafter, the terms of “transcribable protein- or nonprotein-coding sequences”, “transcribable sequences”, “genetic sequences”, and “genes” will be used interchangeably.) These syndromes often have complexed phenotypes and are presented as developmental, autistic, neurological, psychiatric, diabetic, autoimmune, and/or neoplastic diseases [24-28].

To identify the genes that are related to the repair of the poisoned TOP2, as well as their relationships with any specific diseases, we applied the γ H2A.X antibody-dependent chromatin immunoprecipitation-DNA sequencing (ChIP-seq) technology to genome-widely

capture Jurkat cell's genetic sequences that were associated with the repair of the etoposide-induced DSBs. We determined to precipitate the chromatins associated with the DNA repair rather than those only associated with the TOP2 cleavable complexes, for the reason that the poisoned TOP2-induced chromosome aberration often results from the DNA repair that involves DNA sequences far from the TOP2 cleavable complexes [8,29]. As expected, we captured many genetic sequences that are known to contribute to leukemia and autism. However, unexpectedly, a large number of genetic sequences that are related to many other diseases, were also identified. The possible roles of the poisoned TOP2 in the pathogenesis of these diseases are discussed and a hypothetical model is proposed to interpret the complexed features of these diseases.

Results

Poisoned TOP2 has long been known to cause chromosome aberration. However, the vast majority of the involved genes are undefined. Therefore, the only purpose of the current study was to identify the genes that were associated with the repair of the poisoned TOP2 and to find if these genes were related to any specific diseases. It was not our intention to study the chemistry, biochemistry or biology of the poisoned TOP2 or its targeted genes, because a large number of pioneering investigators have done many explicit studies. As a result, we only present the data that are related to the genetic sequences and the associated diseases. We do not show those sophisticated and exquisite figures or data that are generated by the ChIP-seq computer software, the usual way that most ChIP-seq studies do. We would not like

to distract the readers from the main theme of the current study.

2

ChIP captured a number of transcriptable protein- and nonprotein-coding genetic sequences.

We treated the human leukemia Jurkat cells continuously with 100 μ M etoposide, waited for 5.5 hours and then immunoprecipitated the chromatins with the anti- γ H2A.X antibodies. We also immunoprecipitated the chromatins of the cells that were mock-treated with DMSO, which was used to prepare the stock solution of etoposide, to control any endogenous DSBs such as those described by Seo *et al* [5]. We designed our experiments in such ways for the following reasons.

- 1) Unlike other TOP2 poisons such as the bioflavonoids that have multiple cellular effects,
- 2 including antioxidant, pro-oxidant, anti-mutagenic, mutagenic, carcinogenic, and anti-carcinogenic activities, which have not been thoroughly characterized at the molecular level [13], etoposide is well known to poison TOP2 and hence has been used by most investigators to study the poisoned TOP2 [6-8]. We utilized etoposide just as a TOP2-poisoning tool because our only purpose was to identify the genetic sequences associated with the poisoned TOP2 and its repair. We did not intend to characterize any specific TOP2 poisons. Therefore, the principle of the conclusion of the current study should be applicable to other TOP2 poisons such the TOP2-poisoning bioflavonoids, although the details are not likely to be exactly the same because of the disparate efficacy, potency and DNA sequence selectivity among various TOP2 poisons.

2 2) Early studies have immunoprecipitated the TOP2 cleavable complex- or DNA damage-associated chromatins [31,32]. We instead immunoprecipitated the γ H2A.X- or DNA repair-associated chromatins, because the chromosome aberration induced by TOP2 poisons often occur as a result of DNA repair and involves DNA sequences far distant to the DSB [8,29]. Actually, the chromosome aberration is a direct result of the repair (such as homologous recombination and non-homologous end joining) or mis-repair of the poisoned TOP2-induced DSB, rather than the DSB itself. More importantly, the poisoned TOP2 in the TOP2 cleavable complex is quickly degraded through both ubiquitin-dependent and independent proteasome pathways [33-35]. Moreover, γ H2A.X, which only binds to the naked DNA [1-4], is not like to bind to the DNA that is still occupied by the intact or partially degraded TOP2. Therefore, any ChIP-seq studies using the anti-TOP2 antibodies to precipitate the TOP2 cleavable complexes, are likely to underestimate the number of the TOP2 cleavage sites as well as the DNA repair-involved genetic sequences. Indeed, as shown by Dellino *et al.*, the regions enriched with the TOP2B ChIP-seq signals upon the etoposide treatment, do not correlate with those of the γ H2A.X ChIP-seq signals [32].

3) Additional to the DSB, etoposide has also been shown to cause the poisoned TOP2-linked SSB [11,12]. However, we only addressed the TOP2-linked DSB, rather than the SSB, because it is the DSB that is considered to be associated with the TOP2 poison-induced chromosome aberration [13], γ H2A.X formation and cytotoxicity [12]. Moreover, studying the chemistry, biochemistry or biology of the poisoned TOP2 was not the purpose of the current study.

4) We chose γ H2A.X as an indicator of DNA repair, because it is formed at the early stage of DSB repair [1-4]. This way, we avoided the complexed, distinctive cellular outcomes (such as apoptosis, senescence, survival with or without chromosome aberrations, and differentiation). These distinctive cellular outcomes may result from different amounts of DNA damage and/or disparate DNA repair capabilities (i.e. whether the DSB is unrepaired, error-free repaired or error-prone repaired).

5) To capture as many genomic segments that were undergoing repair as possible, we employed 100 μ M etoposide and waited for 5.5 hours before collecting the cells for ChIP-seq. We treated the cells with 100 μ M etoposide because we concerned that a low dose of etoposide might not induce the sufficient amount of a signal that fell into the sensitivity range of the ChIP-seq method [5]; the amount of the signal of any single gene among a heterogeneous cell population is likely to be low. We harvested the cells after the 5.5-hour drug treatment because our kinetic analysis showed that γ H2A.X peaked at around 5 hours following the 100 μ M etoposide treatment (Supporting Information S1 Fig). It should be noted that the etoposide-induced *MLL* fusion seems independent of the dosage of etoposide as a wide range of doses, from 0.14-0.5 μ M, of etoposide can induce the fusion [36,37]. It should also be noted that DNA repair is a kinetically and dynamically changing process [1-4]. Therefore, the γ H2A.X-associated DNA is not likely to be the same throughout the entire DNA repair process. Therefore, the genetic sequences captured by our ChIP only reflected those that were associated with γ H2A.X at 5.5 hour after the etoposide treatment. On the other hand, as we discuss in the following Discussion section, these experimental conditions do not negate our conclusion.

Under the above discussed experimental conditions, a total number of 3707 transcriptable protein- and nonprotein-coding sequences were reproducibly or repeatedly identified by the ChIP-seq (NCBI SRA accession number SRP150381 at <https://www.ncbi.nlm.nih.gov/sra>).

Initially, we investigated these 3707 genetic sequences one by one to find whether they were related to any specific diseases. Three databases, Gene (https://www.ncbi.nlm.nih.gov/gene/), PubMed (https://www.ncbi.nlm.nih.gov/pubmed/) and Database of Genomic Variants (<http://dgv.tcag.ca/dgv/app/home>) were searched. We spent more than 6 months only to screen the first 500 or so chromatin-immunoprecipitated (ChIPed) genetic sequences. We found that at least 50% of them were linked or susceptibly linked to certain known diseases. However, this initial gene-disease correlation study was very laborious and quite inefficient. Additionally, none of the three databases seemed inclusive or updated. More importantly, for a large portion of our ChIPed genetic sequences, scientists have not seemed to reach a consensus on their gene-disease relationships. As a result, we changed our strategies. Instead of investigating every ChIPed genetic sequence, we spent another 6 months or more to focus on the genes of 17 diseases or conditions. The gene-disease relationships of these genes had well been documented in the literature. We chose these 17 diseases also for the reason that since the industrialization, most CNVs and SNPs related to these diseases have occurred *de novo* and the morbidity rates of these diseases have continuously been rising. We contemplated that preventive measures can be introduced to reduce the morbidity rates if the contributing etiologies to the genetic alterations of these diseases are identified.

The poisoned TOP2-induced, γ H2A.X-associated DNA was closely linked to many genetic sequences associated with the CNV-related microdeletion and microduplication syndromes.

Microdeletion and microduplication syndromes refer to a group of diseases that arise from the deletion, duplication or other types of chromosome aberrations of a genomic segment, often leading to CNVs. Clinically, these syndromes may be manifested as developmental, autistic, renal, cardiovascular, neurological, psychiatric, autoimmune, neoplastic, and/or diabetic diseases [24-28]. Our ChIP captured many genetic sequences known to be within the deleted or duplicated chromosome segments of the syndromes. Table 1 lists these CNV-associated syndromes, as well as the involved genes, including their chromosome loci, which were captured by our ChIP. (The genes listed are not necessarily the candidate genes of the diseases and we only list one involved gene although, many times, more genes in the deleted or duplicated genomic segments were pulled down by the ChIP.) As illustrated, the genetic sequences of 66.4% (156 out of 235) of the microdeletion and microduplication syndromes reported by Weise *et al.*, were captured by our ChIP (Table 1) [27]. Therefore, the poisoned TOP2-induced, γ H2A.X-associated DNA was closely linked to many CNV-related microdeletion and microduplication syndromes.

Table 1. Microdeletion and microduplication syndromes and the featured ChIPed genes

Microdeletion and microduplication syndromes reviewed by Weise <i>et al.</i> ²⁷ . Please refer to Weise <i>et al.</i> for genomic coordinates (hg18).	The ChIPed genes within the deleted or duplicated genomic segments; genomic coordinates (kb) (hg18)*
1p36 del**	<i>MORNI</i> ; 2,242-2,312
1p36 del / 1p36 dup***	<i>CHD5</i> ; 6,084-6,162
1p34.1 dup	<i>MAST2</i> ; 46,041-46,274
1p32.2 del	<i>OMAI</i> ; 58,718-58,785
1p21.3 del	<i>LPPR5</i> ; 99,128-99,243
1q21.1 del / 1q21.1 dup	<i>POLR3GL</i> ; 142,400-148,000
thrombocytopenia-absent radius syndrome	<i>POLR4GL</i> ; 142,400-148,000
1q21.1 del / 1q21.1 dup	<i>POLR5GL</i> ; 142,400-148,000
1q24-q25 del	<i>LOC100506023</i> ; 171,470-171,712
1q24.3 del	----
Van der Waude syndrome 1	<i>SYT14</i> ; 208,178-208,404
1q41-q42 del	<i>DISP1</i> ; 221,055-221,245
corpus callosum agenesis del	<i>Clorf101</i> ; 242,691-242,870
2p25.3 dup	<i>TSSC1</i> ; 3,171-3,360
Feingold syndrome	----
hypotonia-cystinuria syndrome	----
holoprosencephaly 2	----

2p21 dup	----
2p16.3 NRXN1 del / 2p16.3 NRXN1 dup	----
2p15–16.1 del	<i>VRK2</i> ; 58,127-58,240
2p14-p15 del	<i>SLC1A4</i> ; 65,069-65,104
2p11.2-p12 del	<i>CTNNA2</i> ; 80,394-80,729
2q11.2 del	<i>FAM178B</i> ; 96,905-96,924
mesomelic dysplasia	<i>AFF3</i> ; 99,530-100,088
2q11.2-q13 del / 2q11.2-q13 dup	<i>MAP4K4</i> ; 101,680-101,877
nephronophthisis 1 / 2q11.2-q13 dup	----
2q13 del / 2q13 dup	<i>TMEM87B</i> ; 112,529-112,593
autism-dyslexia / 2q14.3 dup	<i>CNTNAP5</i> ; 124,499-125,389
Mowat–Wilson syndrome	<i>ZEB2</i> ; 144,858-144,994
2q23.1 del	<i>MBD5</i> ; 148,495-148,987
2q23.3-q24.1 del	<i>GALNTI3</i> ; 154,436-155,018
del 2q24.3 / neonatal epilepsy dup	<i>COBLL1</i> ; 165,244-165,406
synpolydactyly 1 / 2q31.1 dup	----
2q31.2-q32.3 del	<i>UBE2E3</i> ; 181,553-181,636
2q33.1 del	<i>SPATS2L</i> ; 200,881-201,055
brachydactyly-mental retardation syndrome	<i>ANKMY1</i> ; 241,067-241,146
distal 3p deletion	<i>CNTN4</i> ; 2,908-3,074
Von Hippel Lindau disease	----

3p21.31 del	<i>CCDC36</i> ; 49,211-49,270
3p14.1-p13 del	<i>FOXP1</i> ; 71,087-71,715
3p11.1-p12.1 del	----
proximal 3q del syndrome	<i>CBLB</i> ; 106,859-107,070
3q13.31 del	<i>ZBTB20-AS1</i> ; 115,553-115,590
blepharophimosis, ptosis, and epicanthus inversus syndrome	----
Dandy–Walker syndrome	----
3q27.3-q29 del	<i>FGF12</i> ; 193,339-193,609
3q29 del / 3q29 dup	<i>DLG1</i> ; 198,253-198,395
Wolf–Hirschhorn syndrome / 4p16.3 dup	<i>WHSC1</i> ; 1,872-1,920
4p16.1 dup	<i>CLNK</i> ; 10,100-10,295
4p15.3 del	<i>LCORL</i> ; 17,491-17,632
4q21.21-q21.22 del	<i>C4orf22</i> ; 81,475-82,103
4q21 del	----
4q21.2-q21.3 del	----
Parkinson disease 1	<i>SNCA</i> ; 90,864-90,977
Rieger type 1	----
4q32.1-q32.2 triplication syndrome	<i>PDGFC</i> ; 157,902-158,111
Cri–du–Chat syndrome	<i>CTNND2</i> ; 11,024-11,957

Cornelia de Lange syndrome / NIPBL dup	----
spinal muscular atrophy	<i>SMN1</i> ; 70,256-70,284
5q14.3 del	----
5q14.3-q15 del	<i>GPR98</i> ; 89,890-90,495
familial adenomatous polyposis	----
adult-onset autosomal dominant leukodystrophy	<i>LMNB1</i> ; 126,140-126,200
PITX1 del	<i>PCBD2</i> ; 134,268-134,326
5q31.3 del	<i>NDFIP1</i> ; 141,468-141,514
pseudo trisomy 13 syndrome	<i>RANBP17</i> ; 170,221-170,659
5q35.1 del	----
parietal foramina	----
Sotos syndrome / 5q35 dup	<i>SIMC1</i> ; 175,597-175,705
6p del	<i>DUSP22</i> ; 237-296
6p22.3 del	----
adrenal hyperplasia	<i>CYP21A2</i> ; 32,114-32,117
6p21.31 del	<i>PHF1</i> ; 33,486-33,492
6q13–q14 del	<i>KCNQ5</i> ; 73,388-73,965
Prader–Willi like syndrome	----
transient neonatal diabetes mellitus 1	<i>HYMA1</i> ; 144,365-144,371
6q25.2-q25.3 del	<i>ZDHHHC14</i> ; 157,722-158,014
PARK2 del / PARK2 dup	<i>PARK2</i> ; 161,688-163,068

6q27 del anosmia / Chondroma	<i>MLLT4</i> ; 167,970-168,108
Saethre–Chotzen syndrome	----
Greig cephalopolysyndactyly	<i>GLI3</i> ; 41,967-42,243
Williams–Beuren syndrome / 7q11.23 dup	<i>ELN</i> ; 73,080-73,122
WBS-distal deletion	<i>HIP1</i> ; 75,000-75,206
split hand/foot malformation 1	<i>DYNC1H1</i> ; 95,239-95,565
7q22.1-q22.3 del	<i>RELN</i> ; 102,899-103,417
Autism/dyslexia 7q31.1 del	<i>IMMP2L</i> ; 110,090-110,989
speech-language-disorder 1	<i>FOXP2</i> ; 113,513-114,121
holoprosencephaly 3	----
triphalangeal thumb polysyndactyly syndrome	----
Currarino syndrome	----
8p23.1 del / 8p23.1 dup	<i>PPP1R3B</i> ; 9,031-9,045
del 8p21.2	<i>NKX3-1</i> ; 23,592-23,596
del 8p12-p21	<i>DOCK5</i> ; 25,098-25,326
dup 8q11.23	<i>FAM150A</i> ; 53,609-53,640
CHARGE syndrome / 8q12 dup	----
del 8q12.3-q13.2	<i>PDE7A</i> ; 66,789-66,863
mesomelia-synostoses syndrome	----
del 8q21.11	<i>ZFHX4</i> ; 77,756-77,942
nablus mask-like facial syndrome	<i>TRIQC</i> ; 93,964-94,046

del 8q22.2-q22.3	<i>ATP6V1C1</i> ; 104,102-104,154
Langer–Giedion syndrome	----
sex reversal syndrome 4	<i>DOCK8</i> ; 263-455
monosomy 9p syndrome	<i>DOCK8</i> ; 263-455
dup 9q21.11	<i>BANCR</i> ; 71,101-71,111
9q22.3 del / PTCH1 dup	<i>FGD3</i> ; 94,776-94,838
holoprosencephaly 7	----
nail-patella syndrome	<i>LMX1B</i> ; 128,416-128,503
early infantile epileptic encephalopathy 4	----
9q34 del / dup	<i>EXD3</i> ; 139,380-139,437
9q subtelomere deletion	<i>CACNA1B</i> ; 139,892-140,138
hypoparathyroidism, sensorineural deafness, and renal disease	----
Di George syndrome 2	----
10q22-q23 del	<i>NRG3</i> ; 83,627-84,736
juvenile polyposis syndrome	<i>NUTM2A-AS1</i> ; 88,988-89,092
split-hand-foot malformation 3	<i>BTRC</i> ; 103,103-103,307
10q25-q26 del	<i>DOCK1</i> ; 128,584-129,140
Beckwith–Wiedemann syndrome-Silver Russell syndrome (del) / Beckwith–Wiedemann syndrome-Silver Russell syndrome (dup)	----

WAGR syndrome / 11p13 dup	<i>WT1</i> ; 32,365-32,408
Potocki–Shaffer syndrome	<i>ACCS</i> ; 44,044-44,062
spinocerebellar ataxia type 20	<i>MYRF</i> ; 61,279-61,312
11q14.1 del	----
Jacobsen syndrome	<i>CBL</i> ; 118,582-118,684
12p13.31 dup	<i>FOXJ2</i> ; 8,076-8,099
12q14 del	<i>HELB</i> ; 64,982-65,023
nasal speech-hypothyroidism del	<i>CNOT2</i> ; 68,923-69,035
Noonan syndrome 1	<i>PTPN11</i> ; 111,340-111,409
13q12 del / 13q12 dup	----
spastic ataxia Charlevoix–Saguenay	<i>TNFRSF19</i> ; 23,051-23,141
13q12.3-q13.1 del	<i>BRCA2</i> ; 31,787-31,871
retinoblastoma	<i>RBI</i> ; 47,775-47,954
Hirschsprung disease 2	<i>EDNRB</i> ; 77,367-77,390
Holoprosencephaly 5	----
14q11.2 del	----
congenital Rett variant / 14q12 dup	----
14q22-q23 del	<i>LRRC9</i> ; 59,456-59,600
autism spherocytosis del	<i>MTHFD1</i> ; 63,924-63,996
14q32.2 del	<i>EVL</i> ; 99,601-99,680
15q11.2 del / 15q11.2 dup	----

Angelman syndrome Type 1 / 15 dup	<i>ATP10A</i> ; 23,474-23,659
Angelman syndrome Type 2 / 15 dup	<i>ATP10A</i> ; 23,474-23,659
Prader–Willi syndrome Type 1 / 15 dup	<i>ATP10A</i> ; 23,474-23,659
Prader–Willi syndrome Type 2 / 15 dup	<i>ATP10A</i> ; 23,474-23,659
15q13.3 del / 15q13.3 dup	<i>MTMR10</i> ; 29,018-29,071
15q14 del	<i>C15orf41</i> ; 34,720-34,889
deafness and male infertility syndrome	----
15q21 del	<i>USP8</i> ; 48,503-48,580
15q24 del	<i>NPTN-IT1</i> ; 71,646-71,648
15q24 del / 15q24 dup	<i>SIN3A</i> ; 73,448-73,530
orofacial clefting	<i>WDR61</i> ; 76,362-76,378
15q25 del	<i>ALPK3</i> ; 83,160-83,217
15q26.1 del	----
Fryns syndrome	<i>MCTP2</i> ; 92,700-92,747
15q26.2-qter del	<i>ASB7</i> ; 98,960-98,987
ATR-16-syndrome	<i>RHBDF1</i> ; 48-62
tuberous sclerosis del syndrome / tuberous sclerosis dup	----
Rubinstein–Taybi syndrome 1 / Rubinstein–Taybi dup	<i>CREBBP</i> ; 3,715-3,870
16p13.1 del / 16p13.1 dup	<i>NDE1</i> ; 15,651-15,727
16p11.2-p12.2 del / 16p11.2-p12.2 dup	<i>USP31</i> ; 22,980-23,068
16p12.1 del / 16p12.1 dup	<i>CDR2</i> ; 22,264-22,293

16q11.2 distal del / 16q11.2 distal dup	<i>SPNS1</i> ; 28,893-28,903
16p11.2 del / 16p11.2 dup	<i>PAGRI</i> ; 29,735-29,741
16q11.2-q12.1 del	<i>DNAJA2</i> ; 45,546-45,565
16q21-q22 del	----
16q12.1-q12.2 del	<i>ZNF423</i> ; 48,082-48,414
16q24.1 del	<i>CRISPLD2</i> ; 83,411-83,500
FANCA deletion	----
Miller–Dieker syndrome / Miller–Dieker dup	<i>VPS53</i> ; 358-564
17p13.3 del / 17p13.3 dup	<i>RAP1GAP2</i> ; 2,646-2,887
17p13.1 del	<i>DNAH2</i> ; 7,563-7,677
hereditary liability to pressure palsies / Charcot–Marie–Tooth 1A	<i>TVP23C-CDRT4</i> ; 15,280-15,407
Smith–Magenis syndrome / Potocki–Lupski syndrome	<i>RAI1</i> ; 17,525-17,655
neurofibromatosis 1 / NF1 dup	----
17q11.2-q12 del	<i>MYO1D</i> ; 27,843-28,228
17q12a del	<i>ACACA</i> ; 32,516-32,730
renal cysts and diabetes syndrome / 17q12b dup	<i>ACACA</i> ; 32,516-32,730
Van Buchem disease	----
17q21.3 del / 17q21.31 dup	<i>KANSL1</i> ; 41,463-41,625
17q21.31-q21.32 del	<i>MRPL45P2</i> ; 42,882-42,924
17q22-q23.2 del	<i>TOM1L1</i> ; 50,333-50,394

17q23.1– 23.2 dup	<i>BCAS3</i> ; 56,109-56,824
17q24.2-q24.3 del	<i>ARSG</i> ; 63,799-63,928
carney complex syndrome 1	<i>ARSG</i> ; 63,799-63,928
17q24.3 dup	<i>KCNJ16</i> ; 65,582-65,643
holoprosencephaly 4	<i>TGIF1</i> ; 3,443-3,448
proximal 18q del	<i>SLC14A2</i> ; 41,448-41,517
Pitt–Hopkins syndrome	<i>TCF4</i> ; 51,040-51,120
18q22.3-q23 del	<i>ZNF236</i> ; 72,665-72,811
Sotos-like 19p13.2 dup	<i>COL5A3</i> ; 9,931-9,982
19p13.13 del / 19p13.13 dup	<i>DAND5</i> ; 12,941-12,946
19p13.12 del	<i>PKNI</i> ; 14,405-14,443
19p13.11 del	<i>SLC35E1</i> ; 16,521-16,544
19q13.11 del	<i>NUDT19</i> ; 37,874-37,896
Diamond–Blackfan anemia	----
del 20p12.3	----
Alagille syndrome 1	<i>SLX4IP</i> ; 10,363-10,552
20q13.13-q13.2 del	<i>ATP9A</i> ; 49,646-49,818
Albright hereditary osteodystrophy	----
20q13.33 del	<i>DNAJC5</i> ; 61,996-62,037
21q21.1 del	----
21q21.3 dup	<i>JAM2</i> ; 25,933-26,008

platelet disorder	----
Down syndrome	<i>TTC3</i> ; 37,377-37,497
Cat-Eye syndrome	<i>POTEH</i> ; 14,636-14,667
Di George syndrome / 22q11.2 dup	<i>HIRA</i> ; 17,698-17,799
22q11.2 distal del / 22q11.2 distal dup	<i>BCR</i> ; 21,852-21,990
neurofibromatosis 2 del syndrome	<i>NF2</i> ; 28,329-28,409
Phelan–McDermid syndrome / 22q13 dup	----
Leri–Weill dyschondrosteosis	----
X-Linked autism-2	----
Steroid sulphatase deficiency	----
Kallmann syndrome 1	----
MIDAS syndrome	----
Nance–Horan syndrome	----
Xp22.11 del	<i>LOC100873065</i> ; 22,187-23,221
X-linked congenital adrenal hypoplasia / DAX1 dup	----
complex glycerol kinase	----
muscular dystrophy Duchenne	----
Xp11.3 deletion syndrome	----
Goltz syndrome	----
17-beta-hydroxysteroid dehydrogenase X	----

Xq12q13.1 dup	----
X inactivation specific transcript	----
Bruton agammaglobulinemia	----
Xq22.2 del / Pelizaeus–Merzbacher dup	----
Xq22.3q23 del	----
lymphoproliferative syndrome 1	----
X-linked hypopituitarism	----
fragile site mental retardation 1	----
Xq28 del	----
Rett syndrome / MECP2 dup	<i>BCAP31</i> ; 152,619-152,642
sex-determining region Y	----
AZF _a del	----
AZF _b del	----
AZF _b +c del	----
AZF _c del	----

*Genomic coordinates are determined according to the Database of Genomic Variants

(<http://dgv.tcag.ca/dgv/app/home>). ***"Del" represents microdeletion. ***"Dup" represents

microduplication.

We only captured the genes related to 66.4%, rather than all, of the microdeletion and

microduplication syndromes. Genes related to the remaining syndromes were not captured.

Among many possible reasons that may be attributed to, one is that many deletion- and duplication-susceptible DNA segments might be epigenetically silenced due to the Jurkat cell-specific chromatin modification. These DNA segments are not accessible to TOP2. However, investigating the true reason is beyond the purpose and scope of the current study. Much more work is required to find the true attributing reasons which could be related to the ChIP-seq method, data analysis, statistics, cell biology, etc.

We did not either capture all the genes within the depleted and duplicated genomic DNA segment of any single syndrome. We only ChIPed those cells that were treated with etoposide for 5.5 hours. However, the repair of the TOP2-induced DNA break at any specific repair step often only involves a limited section of the DNA segment, we believe that a thorough kinetic and dynamic investigation of the repair process would identify more genetic sequences within the deleted or duplicated genomic segment.

The poisoned TOP2-induced, γ H2A.X-associated DNA was closely linked to many genetic sequences related to autistic spectrum disorders.

To weigh how much the poisoned TOP2 would contribute to any specific types of the microdeletion and microduplication syndromes, we decided to investigate autism or autistic spectrum disorders (ASD) for the following reasons. Genes associated with ASD are often long, implicating that more topological strain is generated during transcription [38]. Moreover, the ASD-related genes are more thoroughly studied. We looked at three online ASD databases, Sfari Gene (<https://gene.sfari.org/database/human-gene/>), Autism Genetic

Database (<http://autism-genetic-db.net/browse/chromosome/y/feature/genes>) and AutismK (http://autismkb.cbi.pku.edu.cn/ranked_gene.php). Of the 1051, 223 and 171 genes respectively reported by the databases, our ChIP captured 353 (33.6%), 70 (31.4%) and 46 (26.9%) of them (Supporting Information S1 Table). Therefore, the poisoned TOP2-induced, γ H2A.X-associated DNA was closely linked to ASD.

The poisoned TOP2-induced, γ H2A.X-associated DNA was closely linked to many genetic sequences related to leukemia and other types of cancers.

As discussed above, additional to the TOP2-poisoning anticancer drugs, the TOP2-poisoning bioflavonoids and pesticides have been linked to the chromosome translocation-related leukemia [7,13,16-18]. We then looked at whether the genes related to leukemia or other types of cancers were immunoprecipitated by our ChIP. Of the 138 genes that cause 132 types of fusions that were reviewed by Zhang and Rowley [39], we captured 55 (39.9%) genes that comprised 101 (76.5%) types of the fusions (Table 2.1 and Supporting Information S2 Table). We next looked at whether the poisoned TOP2 might also contribute to other types of cancers. Of the 82 oncogenes and 63 tumor suppressor genes that were defined according to the stringent criteria [40], our ChIP respectively pulled down 25 (30.5%) and 19 (30.2%) of them (Table 2.2 and Table 2.3). Therefore, the poisoned TOP2-induced, γ H2A.X-associated DNA was closely linked to not only leukemia but also many types of cancers.

2

Table 2. The disease-associated genes captured by the ChIP

<p>1. Leukemia: ChIP captured 56 (40.6%) out of 138 disease-involved genes reported by Zhang and Rowley (DOI: 10.1016/j.dnarep.2006.05.020)</p>
<p><i>ABII, ABL, ACACA, ACS2, AF10, AF1p, AF4p12, AF6, AF6q21, AF9, AF9q34, AFF3 (LAF4), ARG, BCR, CBL, CBP, CDK6, CEV14, CXXC6 (TET1), DDX10, EEFSEC (SELB), ERG, EVI1, FGFR1, FOG2, GAS7, GPHN, GRAF, HCMOGT, HIP1, HLF, LASP1, LPP, MDS1, MLL, MN1, MYH11, MYO18A, Myomegalin, NTRK3, NUP214 (CAN), NUP98, p300, PBX1, PCMI, PDGFRA, PICALM, PMX1 (PRRX1), SEPT 11, SMAP1, TEL, TIF2, TRIM24, TRPS1, ZNF198</i></p>
<p>2. Oncogenes: ChIP captured 25 (30.5%) out of 82 unquestionable oncogenes reviewed by Walker <i>et al.</i> (DOI: 10.1158/0008-5472.CAN-11-2266)</p>
<p><i>ABL1, ABL2, ATF1, BCL11A, BCR, BRAF, CARD11, CBLB, CBLC, CCND3, EGFR, ETV6, EVI1, FGFR1, FGFR2, HMGA2, KIT, MAML2, MITF, MLL (KMT2A), MPL, NTRK1, NUP214, PTPN11, USP6</i></p>
<p>3. Tumor Suppressors: ChIP captured 19 (30.2%) out of 63 unquestionable tumor suppressor genes reviewed by Walker <i>et al.</i> (DOI: 10.1158/0008-5472.CAN-11-2266)</p>
<p><i>BCL11B, BMPR1A, BRCA1, BRCA2, CDK6, CDKN2C, CHEK2, CREB1, CREBBP, EXT2, FBXW7, FOXP1, NF2, NR4A3, NUP98, PALB2, RBI, TSC1, WT1</i></p>
<p>4. Hypertension: ChIP captured 6 (25.0%) out of 15 monogenic hypertension genes and 48 (28.6%) out of 168 GWAS-identified genes reported by Seidel and Scholl (DOI: 10.1152/physiolgenomics.00032.2017)</p>

ADAMTS9, ADCY3, BAT2-BAT5, C3orf38-EPHA3, CACNA1D, CASR, CCDC141, CDC42BPA, CDH13, CERS5, CMIP, COL21A1, EBF1, FGD5, FURIN-FES, FUT9-UFL1, HIPK2, HIVEP3, INSR, KIAA1462, MAP4, MAPK4, MECOM, MKLN1, MRC2, NPR3-C5orf23, NR3C2, PDE3A, PHIP, PLCB1, PLCE1, POC5-SV2C, PRDM16, PYY, SDCCAG8, SH2B3, SLC7A1, SVEP1, TBC1D1-FLJ13197, TCF7L2, TNRC6B, ULK4, WNK1, ZBTB38, ZC3HC1, ZNRF3

5. Obesity: ChIP captured 47 (35.1%) out of 134 GWAS-identified genes reported by Stryjecki *et al.* (DOI: 10.1111/obr.12604)

AGBL4, AKAP6, BRE, C9orf93, CADM2, CALCR, CLIP1, COBLL1, CREB1, DDC, EHBP1, ERBB4, FANCL, FHIT, FOXO3, GALNT10, GBE1, GRID1, HIP1, HS6ST3, HSD17B12, KCNMA1, KCNQ1, LEPR, LMX1B, LRP1B, LRRN6C, MAP2K5, NAV1, NEGR1, NRXN3, NTRK2, PARK2, PGPEP1, PRKCH, RALYL, RPTOR, SDCCAG8, SLC22A3, STXBP6, TCF7L2, TLR4, TNNI3K, TUB, UBE2E3, ZBTB10, ZNF608

6. Stature: ChIP captured 19 (27.5%) out of 69 GWAS-identified stature-associated genes reported by Hirschhorn and Lettre (DOI: 10.1159/000192430)

ACAN, ADAMTS17, ADAMTSL3, BCAS3, BMP6, CABLES1, CDK6, CEP63, DYM, FUBP3, HMGA2, LCORL, NHEJ1, SCMHI, SH3GL3, SPAG17, TRIP11, ZBTB38, ZFHX4

7. Ischemic stroke: ChIP captured 3 (50.0%) out of 6 disease-associated genes reported by Falcone *et al.* (DOI: 10.1016/S1474-4422(14)70029-8) or 6 (46.2%) genes out of 13 disease-associated genes reported by Chauhan and Debette (DOI: 10.1007/s11886-016-0804-z)

CDC5L, HDAC9, NINJ2, SH2B3/ALDH2, PRKCH, ZFHX3

<p>8. Diabetes type I (including a number of autoimmune diseases): ChIP captured 8 (21.6%) out of 37 causal candidate genes reported by Pociot and Lernmark (DOI: 10.1016/S0140-6736(16)30582-7)</p>
<p><i>AFF3, BACH2, CLEC16A, GLIS3, PTPN2, RASGRP1, SH2B3, UBASH3A</i></p>
<p>9. Diabetes type II: ChIP captured 37 (31.9%) genes out of 116 GWAS-identified disease-associated genes reported by Flannick and Florez (DOI: 10.1038/nrg.2016.56)</p>
<p><i>ACSL1, ADAMTS9, ANK1, BCL11A, CDC123-CAMK1D, CENTD2, CMIP, DGKB-TMEM195, DNER, DUSP8, FAF1, GCK, GLIS3, HMGA2, HORMAD2, HSD17B12, ITGB6-RBMS1, KCNQ1, LPP, MACF1, NHEG1, NOTCH2-ADAM30, NRXN3, PAM, PTPRD, RASGRP1, SGCG, ST6GAL1, TBC1D4, TCF7L2, TLE4, TMEM163, UBE2E2, UBE3C, ZFAND3, ZMIZ1, ZZE1</i></p>
<p>10. Coronary artery disease: ChIP captured 16 (28.1%) out of 57 GWAS-identified disease-associated genes reported by Khera and Kathiresan (DOI: 10.1038/nrg.2016.160)</p>
<p><i>BCAS3, EDNRA, FURIN-FES, HDAC9, KIAA1462, KSR2, LPA, MFGE8-ABHD2-SMG6-SRR, PHACTR1, SH2B3, SLC5A3-MRPS6-KCNE2, SVEP1, UBE2Z-GIP-ATP5G1-SNF8, WDR12, ZC3HC1, ZEB2</i></p>
<p>11. Metabolic syndrome (including waist-hip ratio): ChIP captured 17 (30.4%) out of 56 GWAS-identified disease-associated genes reported by Fall and Ingelsson (DOI: 10.1016/j.mce.2012.08.018)</p>
<p><i>ADAMTS9, C12orf51, CADM2, FANCL, HOXC13, ITPR2/SSPN, LRP1B, LRRN6C, MAP2K5, NEGR1, NRXN3, NUDT3, POMC/ADCY3, QPCTL/GIPR, TNNT3K, ZNF608, ZNRF3-KREMEN1</i></p>

<p>12. Aortic aneurysm and aortic dissection: ChIP captured 7 (41.2%) out of 17 disease-associated genes reported by Milewicz <i>et al.</i> (DOI: 10.1146/annurev-med-100415-022956)</p>
<p><i>ELN, FBNI, MYH11, MYLK, PRKG1, TGFB2, TGFBR1</i></p>
<p>13. Osteoporosis: ChIP captured 26 (25.0%) genes out of 104 GWAS-identified disease-associated and monogenic osteoporosis genes reported by Rivadeneira and Makitie (DOI: 10.1016/j.tem.2016.03.006)</p>
<p><i>ANKH, ANO5, CASR, C1CN5, CPED1, DCDC5, ENPP1, ESR1, EYA4, FUBP3, HOXC6, KCNMA1, LEKRI, LRP4, MPP7, NOTCH2, PLOD2, RIN3, RPS6KA5, RUNX2, SEC24D, SMG6, SOX6, SPTBN1, TMEM38B, ZMPSTE24</i></p>
<p>14. Alzheimer disease: ChIP captured 15 (31.3%) out of 48 GWAS-identified disease-associated genes reported by Shen and Jia (DOI: 10.1007/s12264-016-0011-3)</p>
<p><i>ATP6V0A4, ATXN1, CD2AP, CD33, CELF2(CUGBP2), CRI, GOLM1, MTHFD1L, NME8, PCED1B(FAM113B), PICALM, PLXNA4, RGS6, SLC24A4/RIN3, TRPC4AP</i></p>
<p>15. Huntington disease: ChIP captured 6 (50.0%) out of 12 candidate disease modifier SNP genes reported by Correia <i>et al.</i> (DOI: 10.3233/JHD-150169)</p>
<p><i>ATG7, GRIK2, GRIN2A, GRIN2B, HIP1, PPARGC1A</i></p>
<p>16. Parkinson disease: ChIP captured 6 (40.0%) out of 15 causal Parkinson and Parkinson-plus syndrome genes reported by Verstraeten <i>et al.</i> (DOI: 10.1016/j.tig.2015.01.004)</p>
<p><i>SNCA (PARK1 & PARK4), PARK2, LRRK2 (PARK8), ATP13A2 (PARK9), DNAJC6 (PRK19), SYNJI (PARK20)</i></p>

The poisoned TOP2-induced, γ H2A.X-associated DNA was closely linked to a large number of genetic sequences related to many common diseases or conditions.

For the past several decades, the morbidity of many diseases and conditions, such as the cardiovascular and metabolic disorders, has been increased progressively and the onset age has been decreased continuously. On the other hand, the TOP2-poisoning bioflavonoids have been found to be associated with the genetic variations at the early or progressively later stage of a number of these diseases (such as atherosclerosis, cancer, and premature aging) [41]. To investigate whether the poisoned TOP2 might be linked to the genetic alterations of these diseases, we looked at the genes that have been reported to be associated with the other 13 common diseases and conditions, including hypertension, obesity, stature or height, ischemic stroke, type I diabetes, type II diabetes, coronary artery disease, metabolic syndrome (including hip-waist ratio), aortic aneurysm and dissection, osteoporosis, Alzheimer disease, Huntington disease, and Parkinson disease. Table 2.4 to Table 2.16 show that a large percent, ranging from about 25% to more than 50%, of the genes reported to be associated with these diseases or conditions were captured by our ChIP. Interestingly, except the *LPA* gene, all other coronary artery disease-associated genes captured by the ChIP were not related to lipid metabolism [42]. Likewise, most ChIP-captured stroke-associated genes were related to ischemic rather than hemorrhagic stroke [43]. One of the possibilities is that, for the Jurkat cells, the lipid metabolism- and hemorrhagic stroke-associated genes were epigenetically

silenced or not located at the slow DNA replication region.

Additionally, we captured many genes, including *AFF3*, *BACH2*, *CLEC16A*, *GLIS3*, *PTPN2*, *RASGRP1*, *SH2B3*, *UBASH3A*, which contribute to both type I diabetes and other types of autoimmune diseases, supporting the early conclusion that these diseases share the same genetic etiology [44,45].

Overall, our data show that the poisoned TOP2-induced, γ H2A.X-associated DNA was closely linked to many diseases.

2

Discussion

TOP2 poisons are well known to cause various types of chromosome aberrations [6-8,14,15].

However, except a limited number of genes such as the *MLL* gene, most involved genes are not characterized, neither are their relationships with any specific diseases. Therefore, the one and only purpose of the current study is to identify the genetic sequences associated with the poisoned TOP2 and, particularly, its repair. To achieve our goal, we genome-widely

immunoprecipitated the genomic segments that were associated with the γ H2A.X induced by

the TOP2 poison etoposide. As the signal of any γ H2A.X-associated gene among a

heterogeneous cell population is likely to be low, to capture as much amount of the signal as

possible so that it fell into the sensitivity range of the ChIP-seq method, we treated the Jurkat

cells with 100 μ M etoposide, waited for 5.5 hours when the γ H2A.X was approximately at its

peak level (Supporting Information S1 Fig) and then harvested the genomic DNA.

Our repeated ChIP experiments reproducibly captured 3707 transcriptable protein- and nonprotein-coding sequences. These sequences are likely to be TOP2-specific because they include 75.0% of the known topoisomerase-regulated genes essential for neurodevelopment (Supporting Information S3 Table) [46].

The TOP2-poisoning anticancer drugs, bioflavonoids and pesticides have been linked to leukemia and autism [7,13,116-18]. As expected, we did capture many genetic sequences related to leukemia and ASD (Table 2). However, unexpectedly, we pulled down many genetic sequences that belong with 66.4% of the CNV-related microdeletion and microduplication syndromes (Table 1). Moreover, we also identified a large number, ranging from 25% to 50%, of the genetic sequences whose CNVs and/or SNPs were associated with the other 16 common diseases or conditions that we investigated (Table 2). (Oncogenes and tumor suppressor genes belong to one disease, cancer.) These numbers are significant, considering that the human genome contains 154,484 to 323,827 transcriptable protein- and nonprotein-coding sequences [47] whereas we only ChIPed 3707 (1.1%-2.4%) of these sequences (NCBI SRA accession number SRP150381). Therefore, additional to the early reports that have linked TOP2 poisons to leukemia, autism, cardiovascular, and neurological development defects [48-50], our results raise the possibility that the poisoned TOP2 might also be linked to the genetic development of the aforementioned 17 diseases and conditions that we investigated. However, whether this linkage is of cause-effect type is yet to be confirmed.

Like most ChIP-seq studies, we only identified the genes that were related to the repair of

the poisoned TOP2-induced DSB. Our study by no means directly addressed how these genetic sequences might mechanistically cause the CNVs and SNPs; investigating the biochemical or biological mechanism is not the purpose of the current study. However, at least for the *MLL* fusion, several mechanisms have been proposed [13,19-22,51]. While the poisoned TOP2 is considered to directly cause the *MLL* fusion [13], a number of studies have clearly demonstrated that the apoptotic CAD-induced high molecular weight (HMW) DNA cleavage contributes to its formation [19,20]. Indeed, γ H2A.X as well as the non-homologous end joining DNA repair is activated by both poisoned TOP2- and CAD-induced DSBs [3,9,10,52,53]. Therefore, under our experiment conditions (i.e. 5.5 hours of the 100 μ M-etoposide treatment), a portion of the Jurkat cells might have entered apoptosis. As a result, our ChIPed genetic sequences could include those not only due to the initial poisoned TOP2-induced DSBs prior to apoptosis, but also the DSBs resulting from the apoptotic HMW DNA cleavage. However, several pieces of evidence support our conclusion that the ChIPed genetic sequences are poisoned TOP2-specific. The evidence is as follows.

1) The nucleotide-resolution mapping of the TOP2- and CAD-mediated DSBs has found that neither TOP2 cleavage site nor apoptotic DNA cleavage site coincides with the *MLL* translocation hotspot [54].

2) TOP2 cleaves DNA at the anchorage site of the DNA loop domain, generating the apoptotic HMW DNA fragment, independent of the caspase-CAD pathway [55,56].

3) Binding of TOP2 α to CAD greatly enhances TOP2 α 's decatenation activity [57].

4) Simultaneous inhibiting caspase and poisoning TOP2 α abolish the apoptotic

chromosome condensation [57].

These early studies suggest that during apoptosis, TOP2 might work independently or cooperatively with CAD to cleave the looped DNA at the anchorage site, such as the TOP2 cleavage sequence-containing matrix attachment region [20,56,58,59], leading to the apoptotic HMW DNA fragmentation. As suggested by Betti *et al.*, TOP2 poisons might disrupt such roles of TOP2, causing the *MLL* fusion [60]. As a result, regardless whether the poisoned TOP2-induced *MLL* fusion is generated as a result of the initial DNA repair or subsequent apoptosis, it seems that the generation of the *MLL* fusion is related to the poisoned TOP2. Indeed, it has long been known 1) that many DNA damaging agents can induce CAD to cleave DNA within the breakpoint cluster region of the *MLL* gene [61] but, clinically, the therapy-related, *MLL* fusion-associated leukemia is almost always linked to the TOP2-poisoning anticancer drugs [23,62] and 2) that the formation of the *MLL* fusion is independent of the dosage of etoposide, as a wide range of doses, from 0.14-0.5 μ M, of etoposide can induce the fusion [36,37].

TOP2 is an enzyme that resolves the topological strain generated during DNA replication, transcription and other DNA transactions, meaning that TOP2 functions throughout the genome as long as the topological tension interferes with the DNA transactions. Therefore, concerns may be raised regarding 1) whether our ChIPed genetic sequences merely reflect the broad distribution of TOP2 in the genome and 2) whether there is any specificity in terms of the genomic location or disease linkage for the ChIPed genetic sequences over other general genetic sequences. Numerous studies and our findings support that the genomic location of

the ChIPed genetic sequences, as well as their disease linkage, is not random but specific, which are described as follows.

1) The Human Genome Project estimates that the human genome has approximately 154,484 to 323,827 transcribable protein- and nonprotein-coding sequences [47]. The number of the transcribable sequences reproducibly pulled down by our ChIP are much fewer, only 3707 or between 1.1% to 2.4% of the total human sequences.

2) Of the diseases we investigated, we ChIPed 25% to more than 66% of the transcribable sequences that are associated with these diseases (Tables 1 and 2 and Supporting Information S1 and S2 Tables). If the sequences we ChIPed were not specific, average-wise, the number of the ChIPed disease-associated sequences would statically be close to 1.1%-2.4%, the percentage of the ChIPed (3707) over the total transcribable sequences (154,484 to 323,827) in the genome.

3) The intensity of the topological strain, as well as the distribution of TOP2, across the genome or a transcribed and replicated gene is neither identical nor evenly distributed. Usually, more topological strain is accumulated during the transcription of the long genes such as those related to autism [38] and neurodevelopment [46]. Consistently, we ChIPed 33.6% of the autism-related genes (Supporting Information S1 Table) and 75.0% of the known topoisomerase-targeted neurodevelopment genes (Supporting Information S3 Table); both groups of genes are long [38,46].

4) The human genome contains various types of DNA repetitive and other noncanonical sequences [24,25] that may form non-B DNA secondary structures [63]. These noncanonical

DNA sequences or non-B DNA structures are often found at the anchorage site of the looped DNA, such as the matrix or scaffold attachment site, or at the site where DNA transcription or replication is initiated or terminated. At these unique sites, DNA topoisomerases are required to unwind the DNA duplex or resolve the intensified DNA topological problem [6-8,63-70]. Consistently, Yu *et al.* have shown that the TOP2A cleavage sites are not evenly distributed throughout the genome [31]. Moreover, Bariar *et al.* have shown that it is the DNA sequences that drive the illegitimate DNA aberrations [71].

5) Most importantly, numerous studies have confirmed that the chromosome aberration breakpoints associated with CNVs often occur at these non-canonical DNA sequences and non-B DNA structures [63-70,72-74], where TOP2 is concentrated as described above.

6) Our ChIP pulled down 122 pseudogenes, 265 antisense genes and many genes that contain the major common fragile site (such as *ARHGAP15*, *CCSER1*, *CNTNAP2*, *CTNNA1*, *CTNNA3*, *DAB1*, *DLG2*, *DMD*, *DPYD*, *ESRRG*, *FHIT*, *GRID2*, *IL1RAPL1*, *IMMP2L*, *LARGE*, *LRP1B*, *MAD1L1*, *NBEA*, *PARK2*, *PDGFRA*, *PTPRC*, *RORA*, *SDK1*, *THSD7A*, *USH2A*, and *WWOX*) (NCBI SRA accession number SRP150381) [75] and early replicating fragile site (such as *BACH2*, *FOXPI*, *PVT1*, and *PRKCB*) [76]. The chromosome aberration-related DNA segmental duplication and DNA breakage often occur at these unique genetic sequences [29,75,76].

Accordingly, it is not surprising that our ChIPed genetic sequences specifically coincide with those associated with the CNV-related diseases such as the microdeletion and microduplication syndromes.

Additional to the genetic sequences related to the CNVs, we also captured many genetic sequences susceptible to the single nucleotide polymorphisms (SNPs), which is also
 2 consistent with the early findings that many disease-associated SNPs are often closely associated with the CNVs [77].

As a result, our data show that the repair, specifically γ H2A.X activated by the poisoned TOP2 involves many genetic sequences that are linked to a variety of diseases. It should be noted that, we only investigated 17 diseases or conditions. However, our ChIPed genetic sequences involved much more diseases (NCBI SRA accession number SRP150381). For example, we ChIPed the *CFTR* gene whose mutation causes cystic fibrosis. Much more time is required to characterize the gene-disease relationships for all of the ChIPed genetic sequences.

We would like to emphasize that the only direct conclusion obtained from our data is that
 2 the repair of poisoned TOP2 is linked to the genetic sequences that are associated with many diseases. However, whether the linkage is of the etiology-disease (or cause-effect) type remains to be investigated. Currently, only a limited number of genes are known to directly contribute to certain diseases, such as the leukemia-associated *MLL* gene [62]. Nevertheless, the principle of our conclusion (i.e. the repair of poisoned TOP2 is linked to many disease-associated genes) should not be limited only to etoposide. It should also apply to other TOP2 poisons because what we have addressed is the repair of poisoned TOP2. The well-characterized etoposide is only used as a tool to poison TOP2.

Besides our direct conclusion, we also propose an hypothetical model which is describe in

the following paragraphs. It should be noted that unlike the direct conclusion of our study, the
 2 model is extrapolated or deduced from not only our but also others' data. Like most
 hypothetical models, it requires more work to be improved.

To interpret the clinical features of the CNV- and SNP-associated diseases, we propose a
 model according to our and others' data which are as follows.

1) The TOP2-poisoning chemicals, such as the TOP2-poisoning bioflavonoids and
 pesticides, are widely distributed in the environment. Actually, a number of TOP2-poisoning
 bioflavonoids are recommended for the pregnant women as food supplements [50] or are
 contained in the infant formula [78].

2) A large number of previous studies have discovered that the environmental
 TOP2-poisoning chemicals, including the TOP2-poisoning bioflavonoids and pesticides, may
 2 contribute to a variety of human diseases [13-18,41,46,48-50,71,79-87]. More importantly,
 maternal ingestion of the TOP2-poisoning bioflavonoids has been linked to several
 developmental diseases that occur to the fetus and offspring [7,16-18,48-50].

3) Many laboratory studies have found that human and/or mouse monoploid spermatocytes,
 oocytes and diploid embryonic cells are susceptible or related to the TOP2 poison-induced
 chromosome aberration [17,36,37,71,83-86,88].

4) Non-allelic homologous recombination, the most common DNA repair mechanism
 contributing to the CNV-related microdeletion and microduplication syndromes [29], is the
 major DNA repair pathway in meiosis during the maturation of monoploid spermatocytes and

oocytes. These monoploid germ cells have to use the non-allelic homologous recombination
2 to repair the poisoned TOP2-induced DSB because, unlike the diploid cells, the monoploid
cells do not have the allelic genes to conduct the copy-and-paste repair. Therefore, these cells
are more prone to chromosome aberration [29].

5) Many studies have shown that the maternal exposure to the dietary bioflavonoids has a
long-term effect on the offspring later in the adulthood [78].

6) The poisoned TOP2, including that caused by bioflavonoids, is known to cause
chromosome translocation and other types of chromosome aberrations [6-8,13-15] although
the involved genes have not been characterized; the latter is the purpose of the current study.

7) Ames have suggested that the dietary substances are more important than the
environmental and occupational exposures to cause mutation [80,82].

2 8) Our data show that γ H2A.X induced by the poisoned TOP2 is linked to many genetic
sequences whose CNVs and SNPs are related to a variety of diseases (Tables 1 and 2 and
Supporting Information S1 and S2 Tables).

9) Importantly, TOP2 is often located at the noncanonical DNA sequences or non-B DNA
secondary structures which coincide with the breakage sites of various types of chromosome
aberrations and the disease-associated CNVs [6-8,63-70].

10) Any cells, regardless of whether they are cells undergoing meiosis, embryonic cells,
dividing cells or differentiated cells, require TOP2 to resolve the topological problem
generated during the DNA transactions [6-8]. Therefore, they are prone to the mutation

induced by TOP2 poisons.

- 2 11) The previous studies showing the co-occurrence of cancer and many CNV- or chromosome aberration-associated syndromes in children, back our model [89].

To summarize, the poisoned TOP2, on one hand, can cause chromosome aberration and, on the other hand, has been linked to a number of diseases. Moreover, the genomic site (repetitive and other non-canonical sequences) where TOP2 functions, coincides with the site where the chromosome aberration- and CNV-related breakage occurs. Therefore, the disease-associated CNVs and SNPs might be specifically related to the poisoned TOP2, just like the *MLL* fusion that contributes to the pediatric and therapy-related leukemia. Monoploid germ cells are particularly vulnerable because they do not have the allelic genes to conduct the copy-and-paste repair.

- 2 Fig 1 (Please refer to the end of the manuscript for the caption of Fig 1.) shows our model which seems able to explain many clinical features, including mosaicism, two-hit theory, polygenic traits, pleiotropy, and organ- and age-specificity, of both microdeletion and microduplication syndromes and CNV-associated diseases.

The morbidity of many aforementioned diseases, such as ASD and diabetes, have been increased progressively, particularly in the industrialized regions. However, the etiologies for the genetic development of the diseases are elusive. On the other hand, the TOP2-poisoning bioflavonoids are almost ingested daily as food components, additives (in the processed food), supplements, or contaminants (pesticides), often at the high doses that have long been known to cause mutation and genome instability [79,80]. For example, the daily dose of the

2 bioflavonoids recommended by the manufactures can be as high as 1 to 2 grams per day. At
these doses, the plasma levels can be 10 to 20 times higher than those that can induce mouse
MLL fusion [81]. Besides, bioflavonoids at these doses have been linked to a number of
human diseases [7,13,16-18,48-50]. Therefore, a large number of studies have proposed that
the dietary TOP2-poisoning bioflavonoids clinically contribute to the development of many
diseases, including leukemia, autism, developmental diseases, and neurological defects
[7,13,18,41,46,48-50,71,78-87]. Our data that the repair of the poisoned TOP2 involves a
large number of disease-associated genes, support this theory. Moreover, according to our and,
2 particularly, others' data as described above, we extend the theory to include many other
2 diseases and propose that TOP2 poisons, including the TOP2-poisoning bioflavonoids, might
22 also be an important, clinically relevant DNA-mutating etiology for other CNV- and
2 SNP-related diseases. If the hypothesis is confirmed, the morbidity rates of many diseases can
2 be decreased by reducing human's exposure to the TOP2-poisoning substances.

2 It should be noted that only the B- and C-ring dietary bioflavonoids have TOP2 poisoning
2 activities [87]. Moreover, various TOP2-poisoning bioflavonoids exhibit disparate potency and
2 efficacy [71]. Additionally, multiple cellular effects of bioflavonoids have been observed,
2 including antioxidant, pro-oxidant, anti-mutagenic, mutagenic, carcinogenic, and
2 anti-carcinogenic activities among many others [79]. Being antioxidant, antimutagenic and
anticarcinogenic, bioflavonoids also benefit the health for adults [79]. Therefore, more studies,
including modifying the B- and C-ring of the dietary bioflavonoids, might be warranted for
2 their effects on human beings, particularly pregnant women and infants.

Lastly, we hope that any researchers who are interested and are also funded can further the investigation. Afterall, the development disorders, particularly autism, affect many children across the world, particularly in the industrialized regions where the processed food is routinely consumed.

In conclusion, we have found that many genetic sequences associated with the repair of the poisoned TOP2, are the candidates for not only leukemia and ASD but also a variety of other CNV- and SNP-related diseases. Our results raise the possibility that the TOP2-poisoning chemicals such as the dietary bioflavonoids might be involved in the genetic development of many diseases. Additionally, according to our and, particularly, others' data, we propose a yet-to-be-confirmed hypothetical model that seems able to interpret many features of the CNV-related microdeletion and microduplication syndromes (Figure 1). Future cell studies using the TOP2-poisoning bioflavonoids, animal experiments and human epidemiological studies are required to validate the model.

Materials and methods

Cell culture and drug treatment

Human T-cell acute lymphoblastic leukemia Jurkat cells were kindly provided by the Stem Cell Bank of Chinese Academy of Sciences. These Jurkat cells were routinely maintained in the RPMI 1640 media (Gibco). They were mock-treated with 0.1% DMSO (Cat# D2438, Sigma-Aldrich) or treated with 100μM etoposide (Cat# E1383, Sigma-Aldrich) that was

prepared from the 100mM stock solution in DMSO.

Western blot

Cells were collected after a series of time intervals following the continuous treatment of etoposide. Western blot analysis was performed as described previously [30] except that the antibodies against γ H2A.X (Cat# 07-164, Millipore) were used as the primary antibodies at the 1:2000 dilution of manufacture's stock solution.

ChIP-seq

- 2 ChIP was performed using the EZ-ChIP™ Chromatin Immunoprecipitation Kit (Cat# 17-371, Millipore) according to manufacturer's instructions. Both DMSO mock-treated control and etoposide-treated cells were crosslinked with formaldehyde (Cat# F8775, Sigma-Aldrich) at room temperature for 10 minutes. Sonication of the formaldehyde-crosslinked cells was conducted using the Covaris S220 sonicator (Gene Company Limited) at the peak power of 140.0, duty factor of 5.0 and cycle/burst of 200 for 600 seconds. The antibodies against γ H2A.X (Cat# 07-164, Millipore), at the 1:250 dilution from manufacture's stock, were used to immunoprecipitate the etoposide-induced DNA repair complexes. To increase the specificity and reproducibility of the peaks identified by DNA sequencing, 20 to 30 nanograms of DNA from the samples of not only a single ChIP experiment but also a pool of
- 2 8 ChIP experiments were subjected to the DNA sequencing. The sequencing libraries were

prepared following manufacturer's instructions of the SPARK DNA Sample Prep Kit (Cat# SPK0001-V08, Enzymatics). The purified DNA was sequenced paired-end with a read-length of 150 bases on an Illumina HiSeq X Ten sequencer at Shanghai Biotechnology Corporation. The ChIP-seq reads were mapped to the human genome (hg18) with Bowtie2 using the default parameters. The statistically significant peaks ($P < 0.01$) were called using MACS (version 1.4.2) with the default parameters and according to the standard protocols. Both DNA of the DMSO mock-treated cells and input DNA were used for the peak identification of the etoposide-treated samples. Only the peaks that were reproducibly identified from the repeated experiments were annotated to the nearest genes using the NCBI and UCSC databases. The sequencing data were deposited at the NCBI SRA under the accession number SRP150381.

Acknowledgements

Not applicable.

References

1. Rogakou EP, Pilch DR, Orr AH, Ivanova VS, Bonner WM. DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. J Biol Chem 1998 Mar 6;273(10):5858-5868.
2. Kinner A, Wu W, Staudt C, Iliakis G. Gamma-H2AX in recognition and signaling of DNA

- double-strand breaks in the context of chromatin. *Nucleic Acids Res* 2008 Oct;36(17):5678-5694.
3. Mishima M. Chromosomal aberrations, clastogens vs aneugens. *Front Biosci (Schol Ed)* 2017 Jan 1;9:1-16.
4. Georgoulis A, Vorgias CE, Chrousos GP, Rogakou EP. Genome Instability and gammaH2AX. *Int J Mol Sci* 2017 Sep 15;18(9):10.3390/ijms18091979.
5. Seo J, Kim SC, Lee HS, Kim JK, Shon HJ, Salleh NL, et al. Genome-wide profiles of H2AX and gamma-H2AX differentiate endogenous and exogenous DNA damage hotspots in human cells. *Nucleic Acids Res* 2012 Jul;40(13):5965-5974.
- 2 6. Borde V, Duguet M. The mapping of DNA topoisomerase sites in vivo: a tool to enlight the functions of topoisomerases. *Biochimie* 1998 Mar;80(3):223-233.
7. Deweese JE, Osheroff N. The DNA cleavage reaction of topoisomerase II: wolf in sheep's clothing. *Nucleic Acids Res* 2009 Feb;37(3):738-748.
8. Pommier Y, Sun Y, Huang SN, Nitiss JL. Roles of eukaryotic topoisomerases in transcription, replication and genomic stability. *Nat Rev Mol Cell Biol* 2016 Nov;17(11):703-721.
9. Darzynkiewicz Z, Halicka DH, Tanaka T. Cytometric assessment of DNA damage induced by DNA topoisomerase inhibitors. *Methods Mol Biol* 2009;582:145-153.
10. de Campos-Nebel M, Larripa I, Gonzalez-Cid M. Topoisomerase II-mediated DNA damage is differently repaired during the cell cycle by non-homologous end joining and homologous recombination. *PLoS One* 2010 Sep 2;5(9):10.1371/journal.pone.0012541.
11. Bromberg KD, Burgin AB, Osheroff N. A two-drug model for etoposide action against

human topoisomerase IIalpha. J Biol Chem 2003 Feb 28;278(9):7406-7412.

12. Muslimovic A, Nystrom S, Gao Y, Hammarsten O. Numerical analysis of etoposide induced DNA breaks. PLoS One 2009 Jun 10;4(6):e5859.

13. Strick R, Strissel PL, Borgers S, Smith SL, Rowley JD. Dietary bioflavonoids induce cleavage in the MLL gene and may contribute to infant leukemia. Proc Natl Acad Sci U S A 2000 Apr 25;97(9):4790-4795.

14. Ferguson LR, Baguley BC. Topoisomerase II enzymes and mutagenicity. Environ Mol Mutagen 1994;24(4):245-261.

15. Baguley BC, Ferguson LR. Mutagenic properties of topoisomerase-targeted drugs. Biochim Biophys Acta 1998 Oct 1;1400(1-3):213-222.

16. Roman GC. Autism: transient in utero hypothyroxinemia related to maternal flavonoid ingestion during pregnancy and to other environmental antithyroid agents. J Neurol Sci 2007 Nov 15;262(1-2):15-26.

17. Hernandez AF, Menendez P. Linking Pesticide Exposure with Pediatric Leukemia: Potential Underlying Mechanisms. Int J Mol Sci 2016 Mar 29;17(4):461.

18. Bakian AV, VanDerslice JA. Pesticides and autism. BMJ 2019 Mar 20;364:l1149.

19. Stanulla M, Wang J, Chervinsky DS, Thandla S, Aplan PD. DNA cleavage within the MLL breakpoint cluster region is a specific event which occurs as part of higher-order chromatin fragmentation during the initial stages of apoptosis. Mol Cell Biol 1997 Jul;17(7):4070-4079.

20. Hars ES, Lyu YL, Lin CP, Liu LF. Role of apoptotic nuclease caspase-activated DNase in etoposide-induced treatment-related acute myelogenous leukemia. Cancer Res 2006 Sep

- 15;66(18):8975-8979.
21. Gole B, Wiesmuller L. Leukemogenic rearrangements at the mixed lineage leukemia gene (MLL)-multiple rather than a single mechanism. *Front Cell Dev Biol* 2015 Jun 25;3:41.
22. Gothe HJ, Bouwman BAM, Gusmao EG, Piccinno R, Petrosino G, Sayols S, et al. Spatial Chromosome Folding and Active Transcription Drive DNA Fragility and Formation of Oncogenic MLL Translocations. *Mol Cell* 2019 Jul 25;75(2):267-283.e12.
23. Wright RL, Vaughan AT. A systematic description of MLL fusion gene formation. *Crit Rev Oncol Hematol* 2014 Sep;91(3):283-291.
24. Girirajan S, Eichler EE. Phenotypic variability and genetic susceptibility to genomic disorders. *Hum Mol Genet* 2010 Oct 15;19(R2):R176-87.
25. Harel T, Lupski JR. Genomic disorders 20 years on-mechanisms for clinical manifestations. *Clin Genet* 2018 Mar;93(3):439-449.
26. Deak KL, Horn SR, Rehder CW. The evolving picture of microdeletion/microduplication syndromes in the age of microarray analysis: variable expressivity and genomic complexity. *Clin Lab Med* 2011 Dec;31(4):543-64, viii.
27. Weise A, Mrasek K, Klein E, Mulatinho M, Llerena JC,Jr, Hardekopf D, et al. Microdeletion and microduplication syndromes. *J Histochem Cytochem* 2012 May;60(5):346-358.
28. Watson CT, Marques-Bonet T, Sharp AJ, Mefford HC. The genetics of microdeletion and microduplication syndromes: an update. *Annu Rev Genomics Hum Genet* 2014;15:215-244.
29. Carvalho CM, Lupski JR. Mechanisms underlying structural variant formation in genomic disorders. *Nat Rev Genet* 2016 Apr;17(4):224-238.

30. Yang L, Dong F, Yang Q, Yang PF, Wu R, Wu QF, et al. FGF13 Selectively Regulates Heat Nociception by Interacting with Nav1.7. *Neuron* 2017 Feb 22;93(4):806-821.e9.
31. Yu X, Davenport JW, Urtishak KA, Carillo ML, Gosai SJ, Kolaris CP, et al. Genome-wide TOP2A DNA cleavage is biased toward translocated and highly transcribed loci. *Genome Res* 2017 Jul;27(7):1238-1249.
32. Dellino GI, Palluzzi F, Chiariello AM, Piccioni R, Bianco S, Furia L, et al. Release of paused RNA polymerase II at specific loci favors DNA double-strand-break formation and promotes cancer translocations. *Nat Genet* 2019 Jun;51(6):1011-1023.
33. Mao Y, Desai SD, Ting CY, Hwang J, Liu LF. 26 S proteasome-mediated degradation of topoisomerase II cleavable complexes. *J Biol Chem* 2001 Nov 2;276(44):40652-40658.
34. Alchanati I, Teicher C, Cohen G, Shemesh V, Barr HM, Nakache P, et al. The E3 ubiquitin-ligase Bmi1/Ring1A controls the proteasomal degradation of Top2alpha cleavage complex - a potentially new drug target. *PLoS One* 2009 Dec 1;4(12):e8104.
35. Ban Y, Ho CW, Lin RK, Lyu YL, Liu LF. Activation of a novel ubiquitin-independent proteasome pathway when RNA polymerase II encounters a protein roadblock. *Mol Cell Biol* 2013 Oct;33(20):4008-4016.
36. Bueno C, Catalina P, Melen GJ, Montes R, Sanchez L, Ligerio G, et al. Etoposide induces MLL rearrangements and other chromosomal abnormalities in human embryonic stem cells. *Carcinogenesis* 2009 Sep;30(9):1628-1637.
37. Money Penny CG, Shao J, Song Y, Gallagher EP. MLL rearrangements are induced by low doses of etoposide in human fetal hematopoietic stem cells. *Carcinogenesis* 2006 Apr;27(4):874-881.

38. King IF, Yandava CN, Mabb AM, Hsiao JS, Huang HS, Pearson BL, et al.
2 Topoisomerases facilitate transcription of long genes linked to autism. *Nature* 2013 Sep 5;501(7465):58-62.
39. Zhang Y, Rowley JD. Chromatin structural elements and chromosomal translocations in leukemia. *DNA Repair (Amst)* 2006 Sep 8;5(9-10):1282-1297.
40. Walker EJ, Zhang C, Castelo-Branco P, Hawkins C, Wilson W, Zhukova N, et al. Monoallelic expression determines oncogenic progression and outcome in benign and malignant brain tumors. *Cancer Res* 2012 Feb 1;72(3):636-644.
41. Ferguson LR, Philpott M. Nutrition and mutagenesis. *Annu Rev Nutr* 2008;28:313-329.
42. Khera AV, Kathiresan S. Genetics of coronary artery disease: discovery, biology and clinical translation. *Nat Rev Genet* 2017 Jun;18(6):331-344.
- 2 43. Chauhan G, Debette S. Genetic Risk Factors for Ischemic and Hemorrhagic Stroke. *Curr Cardiol Rep* 2016 Dec;18(12):124-016-0804-z.
44. Roizen JD, Bradfield JP, Hakonarson H. Progress in understanding type 1 diabetes through its genetic overlap with other autoimmune diseases. *Curr Diab Rep* 2015 Nov;15(11):102-015-0668-4.
45. Pociot F, Lernmark A. Genetic risk factors for type 1 diabetes. *Lancet* 2016 Jun 4;387(10035):2331-2339.
46. Vokalova L, Durdiakova J, Ostatnikova D. Topoisomerases interlink genetic network underlying autism. *Int J Dev Neurosci* 2015 Dec;47(Pt B):361-368.
47. Salzberg SL. Open questions: How many genes do we have? *BMC Biol* 2018 Aug 20;16(1):94-018-0564-x.

48. Zielinsky P, Piccoli AL,Jr, Manica JL, Nicoloso LH, Menezes H, Busato A, et al. Maternal consumption of polyphenol-rich foods in late pregnancy and fetal ductus arteriosus flow dynamics. J Perinatol 2010 Jan;30(1):17-21.
49. Hahn M, Baierle M, Charao MF, Bubols GB, Gravina FS, Zielinsky P, et al. Polyphenol-rich food general and on pregnancy effects: a review. Drug Chem Toxicol 2017 Jul;40(3):368-374.
50. Barenys M, Masjosthusmann S, Fritsche E. Is Intake of Flavonoid-Based Food Supplements During Pregnancy Safe for the Developing Child? A Literature Review. Curr Drug Targets 2017;18(2):196-231.
- 2 51. Gomez-Herreros F. DNA Double Strand Breaks and Chromosomal Translocations Induced by DNA Topoisomerase II. Front Mol Biosci 2019 Dec 10;6:141.
52. Gomez-Herreros F, Romero-Granados R, Zeng Z, Alvarez-Quilon A, Quintero C, Ju L, et al. TDP2-dependent non-homologous end-joining protects against topoisomerase II-induced DNA breaks and genome instability in cells and in vivo. PLoS Genet 2013;9(3):e1003226.
53. Rogakou EP, Nieves-Neira W, Boon C, Pommier Y, Bonner WM. Initiation of DNA fragmentation during apoptosis induces phosphorylation of H2AX histone at serine 139. J Biol Chem 2000 Mar 31;275(13):9390-9395.
- 2 54. Mirault ME, Boucher P, Tremblay A. Nucleotide-resolution mapping of
2 topoisomerase-mediated and apoptotic DNA strand scissions at or near an MLL translocation
22 hotspot. Am J Hum Genet 2006 Nov;79(5):779-791.
- 2 55. Li TK, Chen AY, Yu C, Mao Y, Wang H, Liu LF. Activation of topoisomerase
2 II-mediated excision of chromosomal DNA loops during oxidative stress. Genes Dev 1999

- 2 Jun 15;13(12):1553-1560.
- 2 56. Solovyan VT, Bezvenyuk ZA, Salminen A, Austin CA, Courtney MJ. The role of
2 topoisomerase II in the excision of DNA loop domains during apoptosis. J Biol Chem 2002
2 Jun 14;277(24):21458-21467.
- 2 57. Durrieu F, Samejima K, Fortune JM, Kandels-Lewis S, Osheroff N, Earnshaw WC. DNA
topoisomerase IIalpha interacts with CAD nuclease and is involved in chromatin
condensation during apoptotic execution. Curr Biol 2000 Jul 27-Aug 10;10(15):923-926.
- 2 58. Khodarev NN, Bennett T, Shearing N, Sokolova I, Koudelik J, Walter S, et al. LINE L1
retrotransposable element is targeted during the initial stages of apoptotic DNA fragmentation.
J Cell Biochem 2000 Sep 7;79(3):486-495.
59. Kantidze OL, Iarovaia OV, Razin SV. Assembly of nuclear matrix-bound protein
complexes involved in non-homologous end joining is induced by inhibition of DNA
topoisomerase II. J Cell Physiol 2006 Jun;207(3):660-667.
60. Betti CJ, Villalobos MJ, Diaz MO, Vaughan AT. Apoptotic stimuli initiate MLL-AF9
translocations that are transcribed in cells capable of division. Cancer Res 2003 Mar
15;63(6):1377-1381.
61. Sim SP, Liu LF. Nucleolytic cleavage of the mixed lineage leukemia breakpoint cluster
2 region during apoptosis. J Biol Chem 2001 Aug 24;276(34):31590-31595.
62. Rashidi A, Fisher SI. Therapy-related acute promyelocytic leukemia: a systematic review.
Med Oncol 2013;30(3):625-013-0625-5. Epub 2013 Jun 15.
63. Wells RD. Discovery of the role of non-B DNA structures in mutagenesis and human
genomic disorders. J Biol Chem 2009 Apr 3;284(14):8997-9009.

64. Bose P, Hermetz KE, Conneely KN, Rudd MK. Tandem repeats and G-rich sequences are enriched at human CNV breakpoints. *PLoS One* 2014 Jul 1;9(7):e101607.
65. Kantidze OL, Razin SV. Chromatin loops, illegitimate recombination, and genome evolution. *Bioessays* 2009 Mar;31(3):278-286.
66. Usdin K, House NC, Freudenreich CH. Repeat instability during DNA repair: Insights from model systems. *Crit Rev Biochem Mol Biol* 2015 Mar-Apr;50(2):142-167.
67. Rothschild G, Basu U. Lingering Questions about Enhancer RNA and Enhancer Transcription-Coupled Genomic Instability. *Trends Genet* 2017 Feb;33(2):143-154.
68. Hall AC, Ostrowski LA, Pietrobon V, Mekhail K. Repetitive DNA loci and their modulation by the non-canonical nucleic acid structures R-loops and G-quadruplexes. *Nucleus* 2017 Mar 4;8(2):162-181.
69. Vasquez KM, Wang G. The yin and yang of repair mechanisms in DNA structure-induced genetic instability. *Mutat Res* 2013 Mar-Apr;743-744:118-131.
70. Canela A, Maman Y, Jung S, Wong N, Callen E, Day A, et al. Genome Organization Drives Chromosome Fragility. *Cell* 2017 Jul 27;170(3):507-521.e18.
71. Bariar B, Vestal CG, Deem B, Goodenow D, Ughetta M, Engledove RW, et al. Bioflavonoids promote stable translocations between MLL-AF9 breakpoint cluster regions independent of normal chromosomal context: Model system to screen environmental risks. *Environ Mol Mutagen* 2019 Mar;60(2):154-167.
72. Cooper GM, Nickerson DA, Eichler EE. Mutational and selective effects on copy-number variants in the human genome. *Nat Genet* 2007 Jul;39(7 Suppl):S22-9.
73. Conrad DF, Pinto D, Redon R, Feuk L, Gokcumen O, Zhang Y, et al. Origins and

- functional impact of copy number variation in the human genome. *Nature* 2010 Apr 1;464(7289):704-712.
74. van Binsbergen E. Origins and breakpoint analyses of copy number variations: up close and personal. *Cytogenet Genome Res* 2011;135(3-4):271-276.
75. Gao G, Smith DI. Very large common fragile site genes and their potential role in cancer development. *Cell Mol Life Sci* 2014 Dec;71(23):4601-4615.
76. Barlow JH, Faryabi RB, Callen E, Wong N, Malhowski A, Chen HT, et al. Identification of early replicating fragile sites that contribute to genome instability. *Cell* 2013 Jan 31;152(3):620-632.
77. McCarroll SA. Extending genome-wide association studies to copy-number variation. *Hum Mol Genet* 2008 Oct 15;17(R2):R135-42.
78. Vanhees K, Vonhogen IG, van Schooten FJ, Godschalk RW. You are what you eat, and so are your children: the impact of micronutrients on the epigenetic programming of offspring. *Cell Mol Life Sci* 2014 Jan;71(2):271-285.
79. Skibola CF, Smith MT. Potential health impacts of excessive flavonoid intake. *Free Radic Biol Med* 2000 Aug;29(3-4):375-383.
80. Ferguson LR. Role of plant polyphenols in genomic stability. *Mutat Res* 2001 Apr 18;475(1-2):89-111.
81. Vanhees K, Coort S, Ruijters EJ, Godschalk RW, van Schooten FJ, Barjesteh van Waalwijk van Doorn-Khosrovani S. Epigenetics: Prenatal exposure to genistein leaves a permanent signature on the hematopoietic lineage. *FASEB J* 2011 Feb;25(2):797-807
82. Ames BN. Identifying environmental chemicals causing mutations and cancer. *Science*

- 1979 May 11;204(4393):587-593.
- 2 83. Ross JA, Potter JD, Reaman GH, Pendergrass TW, Robison LL. Maternal exposure to potential inhibitors of DNA topoisomerase II and infant leukemia (United States): a report from the Children's Cancer Group. *Cancer Causes Control* 1996 Nov;7(6):581-590.
84. Barjesteh van Waalwijk van Doorn-Khosrovani,S., Janssen J, Maas LM, Godschalk RW, Nijhuis JG, van Schooten FJ. Dietary flavonoids induce MLL translocations in primary human CD34+ cells. *Carcinogenesis* 2007 Aug;28(8):1703-1709.
85. Pelkonen O, Terron A, Hernandez AF, Menendez P, Bennekou SH, EFSA WG EPII and its other members. Chemical exposure and infant leukaemia: development of an adverse outcome pathway (AOP) for aetiology and risk assessment research. *Arch Toxicol* 2017 Aug;91(8):2763-2780.
- 2 86. Thys RG, Lehman CE, Pierce LC, Wang YH. Environmental and chemotherapeutic agents induce breakage at genes involved in leukemia-causing gene rearrangements in human hematopoietic stem/progenitor cells. *Mutat Res* 2015 Sep;779:86-95.
87. Bandele OJ, Clawson SJ, Osheroff N. Dietary polyphenols as topoisomerase II poisons: B ring and C ring substituents determine the mechanism of enzyme-mediated DNA cleavage enhancement. *Chem Res Toxicol* 2008 Jun;21(6):1253-1260.
88. Mailhes JB, Marchetti F, Young D, London SN. Numerical and structural chromosome aberrations induced by etoposide (VP16) during oocyte maturation of mice: transmission to one-cell zygotes and damage to dictyate oocytes. *Mutagenesis* 1996 Jul;11(4):357-361.
89. Zimmerman R, Schimmenti L, Spector L. A Catalog of Genetic Syndromes in Childhood Cancer. *Pediatr Blood Cancer* 2015 Dec;62(12):2071-2075.
- 2

Supporting Information Captions

S1 Fig. The kinetics of the etoposide-induced γ H2A.X by the western blot analysis

S1 Table. The ChIP-captured autism-related genes reported by 3 databases

S2 Table. The leukemia chromosome rearrangement-involved genes captured by ChIP

2 **S3 Table. The topoisomerase-regulated neurodevelopment-related genes**
2 **captured by ChIP**

Fig 1. A hypothetical model of the CNV- and SNP-related diseases.

Schematically presented is the maturation process of the parental germ cells and the development of a human being from a fertilized egg (including the two-cell embryo, three germ-cell layer embryo and embryo undergoing organogenesis) to a senior adult. At any developmental stages, the cells are prone to the mutation caused by a TOP2 poison.

The numbered large ovals represent various human organs such as the lungs. The lettered small square boxes represent the genes. Letters A through Z within the small square boxes represent the gene names. Only one set of the alleles are illustrated.

An open square box shows a wild-type gene that is expressed at that specific development stage. A hatched square box represents a mutated gene that is expressed at that specific development stage. Its first appearance indicates that it is initially mutated at that development stage. A closed square box indicates a silenced gene.

Organ 1, 2 and 3 developed from layer I of the three-layer embryo. Organ 5, 6 and 7 developed from layer II of the three-layer embryo. Organ 4 and 8 developed from layer III of the three-layer embryo.

The development stage- and organ-specific exposure to the TOP2 poison determines when and where a gene is mutated. Here, the TOP2 poison-induced genetic mutations occur as follows: Gene A before meiosis (inherited), Genes B and V during meiosis (*de novo*), Genes D (layer I of the three-layer embryo) and X (layer II of the three-layer embryo) at the three-germ layer embryo stage (*de novo*), Genes C, W and Z of Organs 4, 7 and 7 at the junior, middle-aged and senior stages, respectively.

If a gene is mutated during meiosis or at the one-cell embryo stage, all body cells harbor

the mutation (Genes B and V). If the gene of only a portion of the body is mutated after the one-cell embryo stage, mosaicism ensues (Genes D, X, C, W, and Z).

A monogenic disease of Organ 3 develops if a mutated Gene D is sufficient to cause the disease. The polygenic, “two-hit” or modifier-requiring disease of Organ 7 requires all Genes V, W, X, and Z to be mutated accumulatively at several development stages. Pleiotropy occurs when the mutated Gene B causes both diseases of Organ 1 and Organ 2.

Fig 1

bioRxiv preprint doi: <https://doi.org/10.1101/2020.09.03.280669>; this version posted September 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.