

Influence of *ATM*-Mediated DNA Damage Response on Genomic Variation in Human Induced Pluripotent Stem Cells

Junjie Lu,¹ Hu Li,² Anna Baccei,¹ Takayo Sasaki,³ David M. Gilbert,³ and Paul H. Lerou^{1,4}

Genome instability is a potential limitation to the research and therapeutic application of induced pluripotent stem cells (iPSCs). Observed genomic variations reflect the combined activities of DNA damage, cellular DNA damage response (DDR), and selection pressure in culture. To understand the contribution of DDR on the distribution of copy number variations (CNVs) in iPSCs, we mapped CNVs of iPSCs with mutations in the central DDR gene *ATM* onto genome organization landscapes defined by genome-wide replication timing profiles. We show that following reprogramming the early and late replicating genome is differentially affected by CNVs in *ATM*-deficient iPSCs relative to wild-type iPSCs. Specifically, the early replicating regions had increased CNV losses during retroviral (RV) reprogramming. This differential CNV distribution was not present after later passage or after episomal reprogramming. Comparison of different reprogramming methods in the setting of defective DDR reveals unique vulnerability of early replicating open chromatin to RV vectors.

Introduction

DIFFERENT TYPES OF GENETIC ABERRATIONS such as point mutations, small-scale copy number variations (CNVs), and other large-scale chromosomal level changes have been reported in human embryonic stem cells (hESCs) and iPSCs [human pluripotent stem cells (hPSCs)] [1–12]. The presence of these variations is a considerable concern for therapeutic applications of hPSCs, as evidenced by the recent halt of the first human induced pluripotent stem cell (iPSC) clinical study in Japan [13]. While some themes around genomic instability during reprogramming have begun to emerge, our understanding of how genomic aberrations arise in these cells remains limited [14]. Recently, we showed that genome replication timing changes associated with cellular reprogramming shape the CNV landscape in human iPSCs [6]. Replication timing organization is a highly cell-type-specific, spatiotemporally controlled epigenetic property. Replication domains are structural and functional units of the genome with near one to one correspondence to topologically associated domains defined by Hi-C chromosome conformation capture. In addition, replication timing clearly influences genomic mutation rates [15,16]. The relationship between mutation rate and

spatiotemporal organization of the genome underscores the complexity of genome structure and function, and nuclear reprogramming is a powerful platform for studying that relationship [17].

Genome stability, DNA replication, and DNA damage response (DDR) are intrinsically linked to higher-order chromatin organization [18,19]. To study the effect of DDR on genome stability during reprogramming and in pluripotency, we set out to investigate genomic aberrations during factor-based reprogramming when the DDR system has been compromised. Comparisons of these DDR-deficient cells with normal cells could reveal additional properties of genomic variations arising during reprogramming. We focused on the gene *ATM*, in which mutations lead to accumulation of genomic aberrations and result in ataxia telangiectasia (A-T) syndrome. *ATM* is the central gene involved in DDR and repair. Mutations in *ATM* result in defective cell cycle checkpoint activation and a reduced capacity for repair of DNA double-strand breaks (DSBs). iPSCs from A-T patients have been generated by multiple laboratories, but the issue of genomic variation has not been comprehensively investigated [20–24]. We analyze the CNVs of iPSCs derived from A-T patient using high-resolution single-nucleotide polymorphism

¹Department of Pediatric Newborn Medicine and Division of Genetics, Department of Medicine, Brigham and Women's Hospital, Harvard Stem Cell Institute, Harvard Medical School, Boston, Massachusetts.

²Department of Molecular Pharmacology and Experimental Therapeutics, Center for Individualized Medicine, Mayo Clinic, Rochester, Minnesota.

³Department of Biological Sciences, Florida State University, Tallahassee, Florida.

⁴Division of Neonatology, Department of Pediatrics, Massachusetts General Hospital, Boston, Massachusetts.

(SNP) array and discovered differential genome-wide distribution relative to replication timing organization and the effects of integrating versus nonintegrating reprogramming methods.

Materials and Methods

iPSC generation

Dermal fibroblasts from A-T syndrome patients were obtained from the Coriell Institute for Medical Research Cell Repository (Supplementary Fig. S1; Supplementary Data are available online at www.liebertpub.com/scd). These cells and fibroblasts from a healthy control (hFib2) were cultured and reprogrammed as described [25]. Briefly, 100,000 cells were infected with retroviruses prepared from constructs containing GFP-tagged human *OCT4*, *SOX2*, *KLF4*, and *c-MYC* cDNA. On day 4, cells were trypsinized and plated to 10-cm dishes with mitotically inactivated mouse embryonic fibroblasts (iMEFs) as feeder cells. The following day, the medium was changed to hESC medium. Medium was replenished daily. At 20–35 days postinfection, colonies with hESC-like morphology and silenced GFP expression were picked and expanded for further analysis.

Cell culture. Human fibroblast lines were cultured in 10% fetal bovine serum and GlutaMAX in Dulbecco's modified Eagle's medium (DMEM; Lonza). Human iPSCs and ESCs were either cultured on iMEFs in DMEM/F12 supplemented with 20% Knockout Serum Replacement (Gibco), 0.1 mM β -mercaptoethanol (Gibco), GlutaMAX (Gibco), nonessential amino acids (Gibco), and 10 ng/mL FGF2 (Sigma), or in feeder-free conditions on Matrigel in mTeSR medium with 5 \times supplements (STEMCELL Technologies). Human iPSCs were passaged using 50 U/mL type IV collagenase (Gibco) (for feeder culture) or dispase (for feeder-free culture) approximately every 5–7 days.

DNA damage induction

Cells were treated with 25 or 100 ng/mL neocarzinostatin (NCS) or 2 Gy γ -irradiation and fixed at various time points for immunostaining or protein samples were collected for western blotting.

Immunocytochemistry

Samples were washed with phosphate-buffered saline (PBS) and fixed in 4% paraformaldehyde for 20 min at room temperature. After three washes in PBS, cells were permeabilized in 0.2% Triton X-100 in PBST/3% bovine serum albumin (BSA) for 12 min and subsequently washed three times with PBS. After blocking with PBST/3% BSA for 1 h, samples were incubated with primary antibodies diluted in PBST/3% BSA overnight at 4°C. The next day, cells were washed three times with PBST and incubated with secondary antibodies—Alexa Fluor 594 anti-goat IgG or Alexa Fluor 488 anti-rabbit IgG (both from Life Technologies) for 1 h before further washes. Then samples were stained with DAPI and mounted for imaging with a Nikon Ti fluorescence microscope. Primary antibodies used were Oct4 (19857; Abcam), Nanog (21624; Abcam), SSEA-4 (566218; BD Pharmingen), Tra-1-60 (90232; Chemicon), Tra-1-81 (MAB4381; Millipore), phospho-CHK2 (2661; Cell Signaling Technologies), and phospho-H2A.X (05-636; Millipore).

Western blotting

Cells were lysed with RIPA buffer with protease inhibitors and quantified with Pierce BCA assay reagent. Protein samples were separated through sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) and transferred to PVDF membranes, which were then blocked with 5% skim milk in TBST and incubated with primary antibodies overnight at 4°C. Primary antibodies were p53 (SC-126; Santa Cruz), β -actin (612656; BD Biosciences), and ATM (No. 2783; Cell Signaling Technology).

Gene expression array

RNA was isolated with the RNeasy Kit (Qiagen) according to the manufacturer's instruction. The amount of RNA was quantified using a NanoDrop (NanoDrop Technologies) and analyzed with Agilent Bioanalyzer for quality check. For microarray analysis, RNA probes were prepared and hybridized to Human Genome U133 Plus 2.0 oligonucleotide microarrays (Affymetrix) as per the manufacturer's instructions. Arrays were processed by the Coriell Institute Genotyping and Microarray Center. Gene expression levels were normalized with the Robust Multichip Average algorithm. Hierarchical clustering and heat map was performed by means of the Euclidean distance with average linkage method. The calculation was performed by R package. All microarray data have been deposited with the NCBI Gene Expression Omnibus (www.ncbi.nlm.nih.gov/geo/) under accession number GSE78716.

SNP array

Genomic DNA was purified with Qiagen Blood and Tissue DNA Kit. Sample handling and hybridization to Affymetrix SNP 6.0 arrays were performed according to the manufacturer's instructions at the Genotyping and Microarray Center of Coriell Institute for Medical Research. For detecting CNV calls, the Affymetrix Genotyping Console 4.1.1 algorithm was used. CNV locations are based on the human genome assembly of March 2006 (hg18). Samples were normalized to Affymetrix HapMap reference dataset hybridized on the same platform. For CNV calls, regional GC correction, 10-kb cut-off value, and a minimum of 10 markers were used as analysis configurations. All samples passed quality control requirements. All identified CNVs were included, except for CNVs spanning centromeric regions [the average marker distribution within these regions is too large (>40 kb)]. Perl and Python were used for in silico data analysis and CNV data parsing. R language was used for statistical analysis and *P*-value calculations. CNV calls were experimentally validated using quantitative polymerase chain reaction and a computational approach based on replication timing array data as shown previously [6]. Validation rate of iPSC-manifested novel CNVs was about 90%. All SNP array data have been deposited to GSE78715.

Replication timing profiling

Asynchronously cycling iPSC or fibroblast cells are pulse labeled with the nucleotide analog 5-bromo-2-deoxyuridine (BrdU) and sorted into S-phase fractions on the basis of DNA content using flow cytometry. BrdU-labeled DNA from each

fraction is immunoprecipitated, amplified, differentially labeled and cohybridized to a whole-genome comparative genomic hybridization microarray (Roche Nimblegen 6X630k arrays). Data were analyzed as in Lu et al. [6] and deposited to GSE78717.

Results

iPSCs with deficiencies in ATM

Primary dermal fibroblasts from two A-T patients (Supplementary Fig. S1A) were reprogrammed using retroviral (RV) transduction of the standard Yamanaka factors, as we previously reported [6]. The disease cells were obtained from the Coriell Institute for Medical Research Repository. The exact mutations in the two A-T fibroblast lines (ATF1 and ATF2) are not clear due to the large size of the *ATM* gene, but both patients are clinically affected. All iPSC lines express pluripotency markers (Supplementary Fig. S1B) and yielded teratomas comprised of tissues of all three germ layers (Supplementary Fig. S1C).

To confirm that A-T disease iPSCs (A-T iPSCs) retained the *ATM* deficiency phenotype, DSB's were induced using two different approaches—NCS and γ -irradiation. Immunofluorescence and western blotting confirmed that in *ATM*-deficient fibroblasts and the resultant iPSCs, CHK2 phosphorylation is dramatically reduced along with less significant upregulation of p53 (Fig. 1 and Supplementary Fig. S1D). Collectively, these data demonstrate that the A-T iPSCs maintain their original *ATM* deficiency phenotype.

iPSC lines, analyzed at various passages, have a global gene expression pattern indistinguishable from a well-characterized hESC line, CHB-8 (Supplementary Fig. S2). We were not able to detect any consistently up or down-regulated genes or pathways in pluripotent state between A-T iPSCs and iPSCs generated from healthy donor fibroblasts (WT iPSCs). These data show that *ATM* deficiency does not affect global gene expression in the pluripotent state, similar to previous reports [21–24].

Because *ATM* is involved in cell cycle checkpoints, we examined the cell cycle distribution of A-T iPSCs, WT iPSCs,

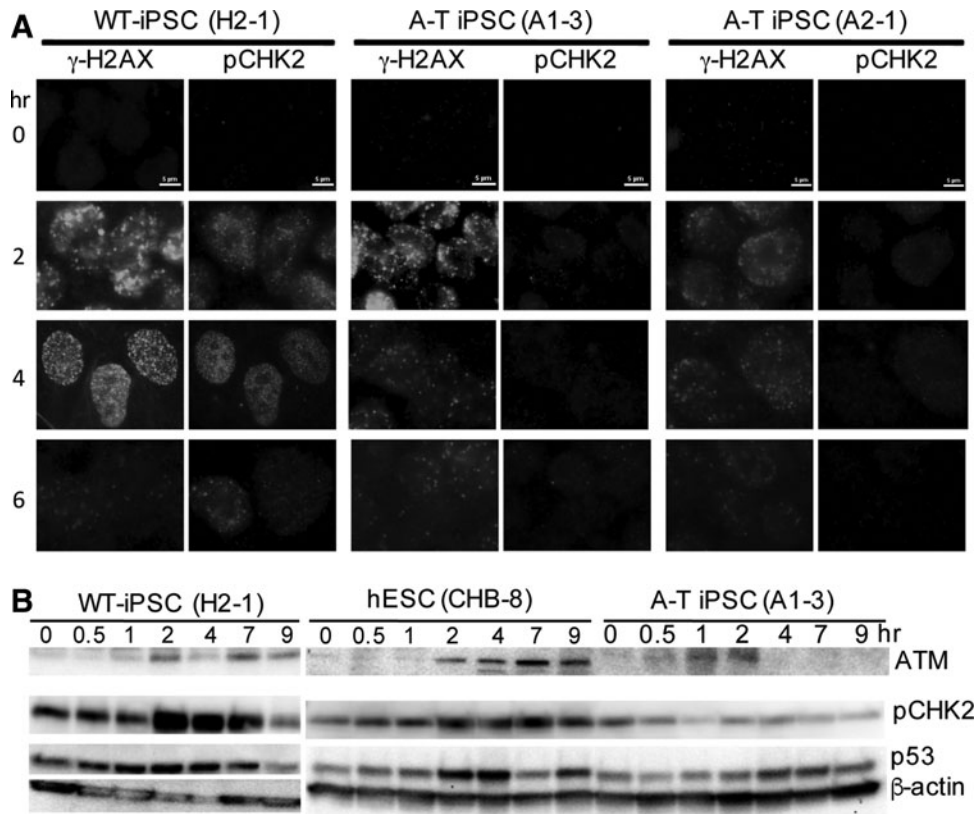


FIG. 1. A-T iPSCs have defective DDR. **(A)** Deficient CHK2 phosphorylation in A-T iPSCs in response to DSBs induced by NCS. A1-3 and A2-1 are A-T iPSC lines generated from A-T fibroblast lines ATF1 and ATF2, respectively. A-T and WT iPSCs (line name: H2-1) were treated with 100 ng/ μ L NCS and fixed for staining at the indicated time. γ -H2AX is a marker of DNA DSBs. CHK2, a major target of ATM, is quickly phosphorylated in WT iPSCs as part of DDR process. In A-T iPSC lines, due to defective ATM function, CHK2 phosphorylation was much weaker and was cleared at a slower speed compared to WT. Scale bars: 5 μ m. **(B)** Deficient CHK2 phosphorylation in A-T iPSCs in response to DSBs induced by γ -irradiation. A-T, WT iPSCs and hESCs (CHB-8) were treated with 2 Gy γ -irradiation and collected at different time points after treatment. Western blots confirmed the lack of ATM, deficient CHK2 phosphorylation, and slower p53 activation in A-T iPSCs. Note the relative similar amount of p-CHK2 and p53 at 0 h. A-T, ataxia telangiectasia; DDR, DNA damage response; DSBs, double-strand breaks; hESCs, human embryonic stem cells; iPSCs, induced pluripotent stem cells; NCS, neocarzinostatin.

and WT hESC lines. We could not observe any obvious difference in the cell cycle profiles among these cell lines (Supplementary Fig. S1E). Under standard culture condition, these cells maintain a similar passage ratio every 6–7 days and could be maintained in culture for over 50 passages.

CNV detection and characterization

We identified CNVs by high-resolution SNP array on six RV reprogrammed A-T iPSC lines and their parental fibroblasts (Supplementary Tables S1 and S2), including low-passage (immediately after reprogramming, passage 5–8) and high-passage (passage 19–23) iPSCs. We also expanded two previously generated WT iPSC lines [6] to similar passages and analyzed their CNVs (Supplementary Tables S1 and S2).

Similar to what we reported previously, the majority of CNVs in iPSCs was novel—not detected in parental fibroblasts [6]. Novel CNVs detected only in low passage cells most likely reflect genomic aberrations that arise or are selected for during the reprogramming process, whereas novel CNVs only detected in high-passage cells reflect the result of prolonged culture in pluripotent state. An initial characterization of the CNVs demonstrated substantial variability in the number of CNVs across the various lines and passages (Fig. 2A and Supplementary Table S1), consistent with what has been previously reported in similar analyses [4–6,26]. Most iPSC CNVs were novel—only a minority was also present in parental fibroblast (Fig. 2B). Of the novel CNVs detected in low passage iPSCs, most (~70%) were unique and no longer detected at high passage (Fig. 2B; 73.8% ± 20.9% for WT iPSCs; 69.1% ± 21.4% for A-T iPSCs). However, compared to WT iPSCs, A-T iPSCs

tended to have a higher number of unique novel CNV losses at higher passage (Fig. 2B; 57.4% ± 16.6% for WT and 81.5% ± 11.5% for A-T). These data demonstrate that CNVs, which appear immediately following reprogramming, are highly dynamic and that more de novo CNVs accumulate in A-T iPSCs during prolonged pluripotent culture, likely due to the DDR defects.

We failed to detect any consistent CNV losses or gains, but we did detect novel CNV gains in multiple A-T iPSC lines in the immunoglobulin (Ig) regions, in particular IgLV locus on chr 22, which was amplified in two of the three A-T iPSC lines from ATF1 after reprogramming (Supplementary Fig. S2B). Other Ig loci (IgKV locus on chr 2 and IgH locus on chr 14) were also affected in some lines, although these were inherited from parental fibroblasts. No genomic level studies have been done in A-T patients to suggest that mutations in this locus could be linked to ATM deficiency. However, the findings that Ig loci are CNV hotspots in A-T iPSCs, but not in WT iPSCs, could potentially be connected with immunological dysfunction seen in A-T patients [27].

ATM deficiency does not alter CNV location within replication domains and replication timing transition regions

WT and ATM-deficient cells undergoing reprogramming will experience a similar degree of genomic stress due to replicative stress, metabolic changes, and reactive oxygen species production associated with reprogramming [28]. Without functional ATM, some of the damage associated with these stresses will not be properly repaired in A-T iPSCs and comparison to WT will reveal the influence of ATM-mediated

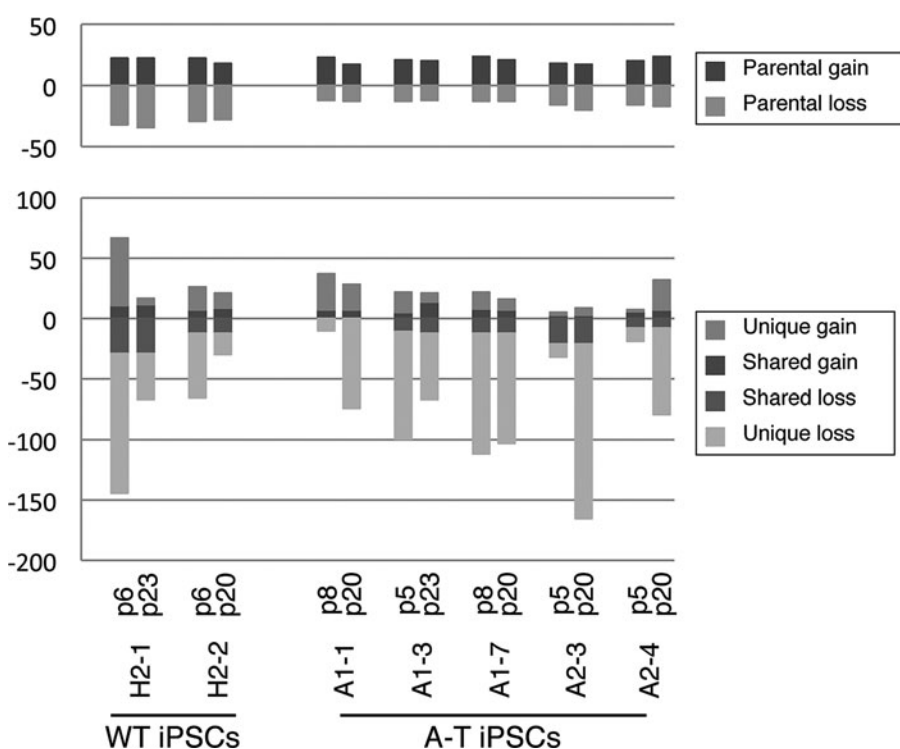


FIG. 2. CNV dynamics during reprogramming and pluripotent culture. CNVs change dynamically between low and high passages. Numbers of CNVs for WT and A-T iPSC lines analyzed by SNP arrays at both low and high passages are plotted. CNV gains and losses in iPSCs are either retained from parental fibroblasts (*upper panel*) or occurred de novo (*lower panel*) during reprogramming (low passage) or pluripotent culture (high passage). A fraction of novel CNVs are present in both low- and high-passage iPSCs (shared CNVs, *darker colors*). The majority of novel CNVs is unique—found at either low passage or high passage, but not both (unique CNVs, *lighter colors*). A1-1, A1-3, and A1-7 are A-T iPSC lines generated from A-T fibroblast line ATF1. A2-3, A2-4 were generated from ATF2. H2-1 and H2-2 are iPSC lines generated from a healthy male donor fibroblast line. For detailed characteristics of CNVs, Supplementary Tables S1 and S2. CNV, copy number variation; SNP, single-nucleotide polymorphism.

DDR on the genome aberration landscape. We generated A-T iPSC-specific replication timing profiles from a representative iPSC line of each A-T patient. Because the two A-T iPSC replication timing profiles are essentially identical (Pearson correlation 0.94), we combined them into one profile for further analyses (Fig. 3A). Of note, the replication timing profile of A-T iPSCs was also similar to that of WT iPSCs (Pearson correlation 0.87) (Supplementary Fig. S3), consistent with a recent report, where ATM knockdown did not change genome-wide chromatin interaction patterns in mouse pro-B cells [29].

Mapping novel CNVs of A-T iPSCs to this replication timing profile, we found that CNVs are contained within replication domains or constant timing regions, CTRs, corresponding to several tandem replication domain that replicate at the same time (Fig. 3B) [30]. We also did not find an enrichment of CNVs in timing transition regions of iPSCs or fibroblasts (Fig. 3C), which are thought to be particularly vulnerable to DNA damage possibly due to the lack of active replication origins [31,32]. These data are similar to what we found with CNVs in WT iPSCs following reprogramming and demonstrate that the distribution of CNVs is related to higher-order replication domain organization [6].

A-T iPSCs have increased CNV losses in early replicating genome

Next, we analyzed novel CNV distribution following reprogramming as a function of replication timing in A-T iPSCs. At low passage, following reprogramming, WT and

A-T iPSCs have similar numbers of gains, but we found a slight predominance of gains in the late replicating domains (Fig. 4A, orange), in contrast to WT, where gains enriched in early replicating domains (Fig. 4A, purple). A-T iPSC had a much higher number of losses compared to WT and they were distributed equally across early and late replicating domains (Fig. 4B, orange). This finding is also in contrast to WT, where CNV losses were strongly enriched in late replicating domains (Fig. 4B, purple). To exclude the potential contribution of the different genetic background (instead of the contribution of ATM deficiency), we separately analyzed CNVs from lines derived from each A-T patient and found similar trends in both (Supplementary Fig. S4A, B). Overall, A-T-iPSCs had a higher load of CNVs in early replicating genome in contrast to WT cells. These results suggest a higher vulnerability of early replication domains to genomic stress during reprogramming.

Reprogramming methods generate distinct genomic stress that shapes the CNV landscape

To test whether genomic stress during reprogramming is related to the integrative nature of the RV vectors, we generated additional A-T (from one of the patient fibroblast cells) and WT iPSC lines using episomal (EP) vectors and profiled CNVs in low-passage cells (Supplementary Tables S1 and S2). The CNV distribution of low-passage EP A-T iPSCs was similar to EP WT iPSCs (Fig. 4C, D), both similar to RV WT iPSCs. Therefore, the differential distribution of CNVs

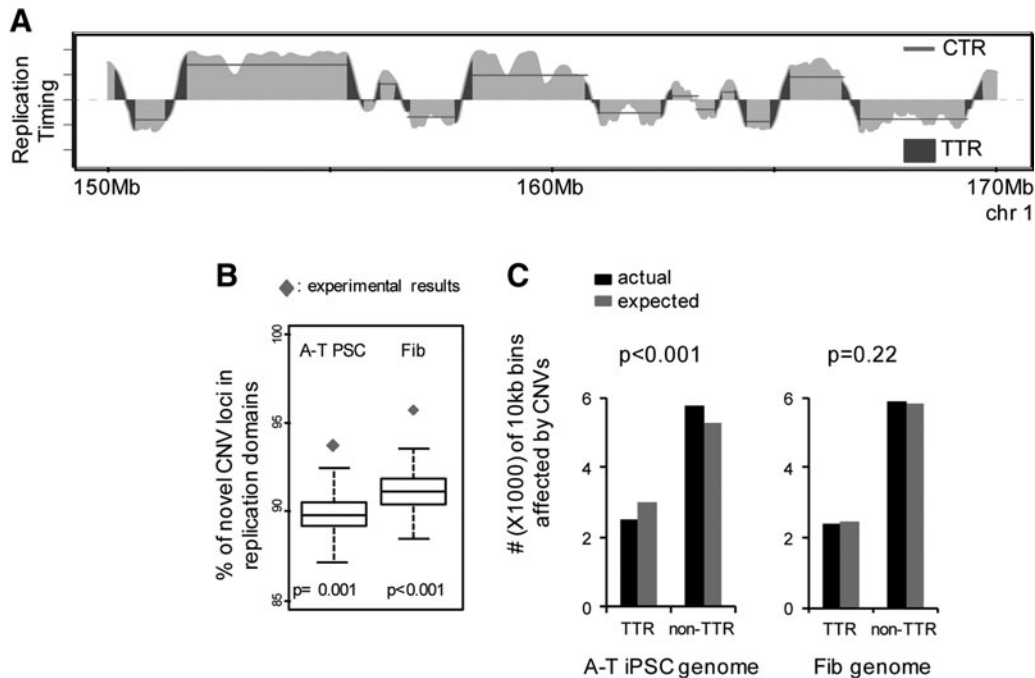


FIG. 3. A-T iPSC CNVs reside in replication domains and are excluded from replication TTRs. **(A)** An exemplary replication timing profile, demonstrating CTRs (could consist of multiple replication domains) and TTRs. **(B)** Permutation analysis demonstrates that CNVs are significantly more likely to reside within individual replication domains than expected by chance in both iPSCs and parental fibroblasts (fib). Boxes display the percentages obtained from 1,000 simulations, and gray diamonds represent the actual percentages. **(C)** TTRs are not enriched within CNVs. The actual amount of TTR and non-TTRs affected by CNVs across the genome (black), compared to expected amount of TTR and non-TTRs affected by random distribution (gray), is shown with *P*-values (Fisher's exact test for 2×2 contingency tables) above each plot. TTRs, timing transition regions.

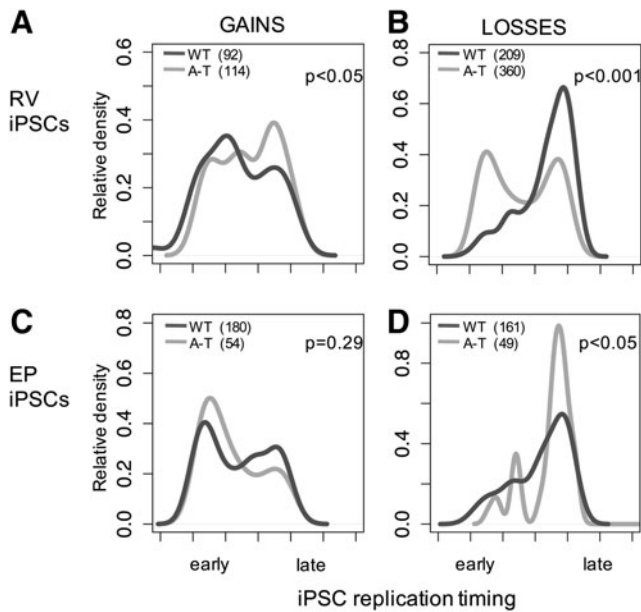


FIG. 4. ATM deficiency reveals unique genomic vulnerability of early replicating domains during RV reprogramming. Kernel density plots show the replication timing distribution of CNV gains and losses detected in WT (*black*) and A-T (*gray*) iPSCs derived using RV or EP vectors. Numbers of CNVs in parentheses. *P*-values calculated using K-S test. **(A)** CNV gains in low-passage (p4–8) A-T iPSCs generated using RV reprogramming (*left panel*) show a slight predominance in late replicating genome regions compared to WT iPSCs. **(B)** CNV losses in RV AT-iPSCs are more enriched in early replicating genome compared to RV WT-iPSCs. **(C, D)** CNV gains and losses in EP-reprogrammed low-passage A-T and WT iPSCs showed overall similar relative enrichment in early and late replicating genome, respectively. Note: comparisons of A-T iPSC CNVs between RV and EP reprogramming are also statistically significant ($P < 0.01$ for gains, *gray lines* between **A** and **C**; $P < 0.001$ for losses, *gray lines* between **B** and **D**, K-S tests), supporting the reprogramming methods related effect on CNV distribution. EP, episomal; K-S, Kolmogorov–Smirnov; RV, retroviral.

in low-passage A-T iPSCs is only seen during RV reprogramming (Fig. 4A, B), suggesting that integrative vector-mediated reprogramming causes more genomic stress in early replicating regions, a trend that is only uncovered in the absence of normal DDR.

CNV distribution in high-passage A-T iPSCs resembles WT iPSCs

To better understand how replicative stress and the pluripotent genome architecture interact to shape the CNV landscape during regular maintenance culture, we compared CNVs detected in low- and high-passage WT and A-T iPSCs. While A-T iPSCs accrued more CNVs during continued culture, especially losses, at higher passage, both A-T and WT iPSCs had a similar distribution of gains and losses relative to low-passage WT iPSCs (Supplementary Fig. S4C, D): gains enriching in early replicating regions and losses enriching in late replicating regions. The increased vulnerability of early replicating genome to CNVs (Fig. 4A, B)

was not seen at the higher passage. These data further support that early replicating regions of the genome are uniquely susceptible to CNVs caused by genomic stress during reprogramming.

Discussion

Various sources of genomic stress, DDR pathways, cellular proliferation, and selective pressure all converge to generate the CNV landscape in cells. These processes operate on the platform of precise, spatiotemporally regulated, cell-type-specific higher-order chromatin organization [17]. Studying the dynamics of these processes during reprogramming and pluripotent culture is inherently difficult because they are so intertwined. By tracking CNV distribution as a function of replication timing, passage, and reprogramming method in the setting of normal (WT) and impaired (A-T) DDR, we discovered important insights about the origin of genomic variations in iPSCs. Our previous study on CNVs in WT iPSCs identified patterns of genomic instability relative to genome organization and reorganization during reprogramming. In this study, we show that genomic stress associated with integrative vector-based reprogramming differentially affects early replicating regions during the reprogramming process. This genomic stress is revealed when ATM function is impaired and is distinct from the genomic stress associated with maintenance culture of iPSCs.

Because DNA DSBs contribute to CNVs and A-T iPSCs have less efficient DSB repair, the distribution of CNVs in these cells reveals the landscape of DSB lesions before repair. Moreover, CNVs detected at low passage and not present in parental fibroblast are considered novel and likely arose during reprogramming. Our analyses demonstrate that novel CNV gains and losses are differentially distributed when ATM-mediated repair is impaired during reprogramming. Specifically, early replication genome is more susceptible to CNV losses (Fig. 4B). Moreover, by comparing CNVs in low and high passage to the parental fibroblasts, we gain insight to the role of DSB repair contributing to CNVs in reprogramming compared to pluripotent culture. In higher passage A-T iPSCs, where novel CNVs reflect genomic instability due to pluripotent culture rather than reprogramming, CNV distributions are similar to WT iPSCs. These findings suggest that the propensity for DSBs to generate CNVs during reprogramming is different compared to pluripotent culture.

Finally, we compared sister iPSCs reprogrammed by different methods of delivering the reprogramming factors. By removing ATM-mediated DDR, we revealed that early replicating/open chromatin domains are more vulnerable to genomic stress in retrovirus-mediated reprogramming compared to nonintegrating EP factors. RV integration happens during or after DNA replication [33–35] and could preferably target early replicating or open chromatin regions, although a genome-wide analysis of RV integration sites relative to higher-order chromatin organization has not been done. Although not tested here, nonrandom genomic stress could be associated with lentiviral vectors and piggyback transposons and, therefore, the genomic aberration effects of different reprogramming methods remain an important safety consideration for iPSCs. Interestingly, a recent report also showed reprogramming method-related differences in genetic variants [12]. Fortunately, the use of integrative vectors for reprogramming has become less popular as other

delivery methods such as sendai viral vectors and mRNAs have become more efficient and readily available. However, the reprogramming factor delivery method is not the only source of genomic stress. Replicative stress during nuclear reprogramming of cells contributes to genomic instability as shown in a drug-inducible system [36].

A significant portion of the genome changes replication timing as a result of nuclear reprogramming. We have previously reported that how the genome's replication timing is changed during reprogramming influences CNV distribution. Specifically, CNV gains accumulate in regions of the genome that change to earlier replication during the reprogramming process. Moreover, this differential distribution of CNVs occurs irrespective of the reprogramming method used. Interestingly, this finding holds true for A-T iPSCs—in both RV and EP reprogramming CNV gains enriched in the genome that changes to earlier replication during reprogramming (Supplementary Fig. S5). These results provide further proof that the increased CNV losses found in the early replicating genome in RV A-T iPSCs, but not in EPI A-T iPSCs, reflect unique genomic stress associated with RV reprogramming that is uncovered by A-T deficiency.

Our finding that iPSC CNV gains and losses in CNVs have differential distribution in open and closed chromatin provide further support for differential repair efficiency of the DDR pathways in different chromatin compartments [37,38]. ATM is more efficient at repairing DSBs, especially losses, in euchromatin (in general early replicating) than in heterochromatin (usually late replicating) [38]. Therefore, during RV reprogramming of WT cells, most losses accumulate in late replicating regions, whereas gains in early regions. However, when ATM-mediated repair is not functioning, these differences are eliminated and the result is a more even distribution of gains and losses across the early and late replicating regions, as demonstrated in low-passage A-T iPSCs.

The experiments involving ATM deficiency illustrate how our understanding of the complex interactions between the DDR and pluripotency network can be informed by surveying the mutational landscape in the context of genome spatial organization. Interestingly, Supek and Lehner recently reported a similar relationship between DNA mismatch repair genes and differential mutation rates across the human genome [39]. Additional studies with more targeted interrogation of other DDR network factors are warranted to better understand how these processes interact with nuclear architecture and cell fate change to result in genomic aberrations and how these aberrations impact the utility of iPSCs as platforms to study the disease and develop new therapies.

Acknowledgments

The authors thank Jeanne Carroll and Ruggero Spadafora for critical reading of the article. Norman Gerry at the Coriell Institute for technical assistance. This work was supported by the NICHD grant R00HD061981 to P.H.L. H.L. is supported by Mayo Clinic Center for Individualized Medicine and Mayo Clinic Center for Regenerative Medicine. D.M.G. is supported by R21CA161666 and P01GM085354.

Author Disclosure Statement

No competing financial interests exist.

References

- Bai Q, J-M Ramirez, F Becker, V Pantesco, T Lavabre-Bertrand, O Hovatta, J-M Lemaitre, F Pellestor and J De Vos. (2015). Temporal analysis of genome alterations induced by single-cell passaging in human embryonic stem cells. *Stem Cells Dev* 24:653–662.
- González F, D Georgieva, F Vanoli, Z-D Shi, M Stadtfeld, T Ludwig, M Jasin and D Huangfu. (2013). Homologous recombination DNA repair genes play a critical role in reprogramming to a pluripotent state. *Cell Rep* 3:651–660.
- Gore A, Z Li, H-L Fung, JE Young, S Agarwal, J Antosiewicz-Bourget, I Canto, A Giorgetti, MA Israel, et al. (2011). Somatic coding mutations in human induced pluripotent stem cells. *Nature* 471:63–67.
- Hussein SM, NN Batada, S Vuoristo, RW Ching, R Autio, E Närvä, S Ng, M Sourour, R Hämäläinen, et al. (2011). Copy number variation and selection during reprogramming to pluripotency. *Nature* 471:58–62.
- Laurent LC, I Ulitsky, I Slavin, H Tran, A Schork, R Morey, C Lynch, JV Harness, S Lee, et al. (2011). Dynamic changes in the copy number of pluripotency and cell proliferation genes in human ESCs and iPSCs during reprogramming and time in culture. *Cell Stem Cell* 8: 106–118.
- Lu J, H Li, M Hu, T Sasaki, A Baccei, DM Gilbert, JS Liu, JJ Collins and PH Lerou. (2014). The distribution of genomic variations in human iPSCs is related to replication-timing reorganization during reprogramming. *Cell Rep* 7: 70–78.
- Ma N, Y Shan, B Liao, G Kong, C Wang, K Huang, H Zhang, X Cai, S Chen, et al. (2015). Factor-induced reprogramming and zinc finger nuclease-aided gene targeting cause different genome instability in β -thalassemia induced pluripotent stem cells (iPSCs). *J Biol Chem* 290: 12079–12089.
- Martins-Taylor K, BS Nisler, SM Taapken, T Compton, L Crandall, KD Montgomery, M Lalonde and R-H Xu. (2011). Recurrent copy number variations in human induced pluripotent stem cells. *Nat Biotechnol* 29:488–491.
- Mayshar Y, U Ben-David, N Lavon, J-C Biancotti, B Yankir, AT Clark, K Plath, WE Lowry and N Benvenisty. (2010). Identification and classification of chromosomal aberrations in human induced pluripotent stem cells. *Cell Stem Cell* 7:521–531.
- Schlaeger TM, L Daheron, TR Brickler, S Entwisle, K Chan, A Cianci, A DeVine, A Ettenger, K Fitzgerald, et al. (2015). A comparison of non-integrating reprogramming methods. *Nat Biotechnol* 33:58–63.
- Wu H, KJ Kim, K Mehta, S Paxia, A Sundstrom, T Anantharaman, AI Kuraishy, T Doan, J Ghosh, et al. (2008). Copy number variant analysis of human embryonic stem cells. *Stem Cells* 26:1484–1489.
- Bhutani K, KL Nazor, R Williams, H Tran, H Dai, Ž Džakula, EH Cho, AWC Pang, M Rao, et al. (2016). Whole-genome mutational burden analysis of three pluripotency induction methods. *Nat Commun* 7:10536.
- Garber K. (2015). RIKEN suspends first clinical trial involving induced pluripotent stem cells. *Nat Biotechnol* 33: 890–891.
- Qi S, Z Fang, D Wang, P Menendez, K Yao and J Ji. (2015). Concise review: induced pluripotency by defined factors: prey of oxidative stress. *Stem Cells* 33:1371–1376.
- Koren A, P Polak, J Nemes, JJ Michaelson, J Sebat, SR Sunyaev and SA McCarroll. (2012). Differential relation-

- ship of DNA replication timing to different forms of human mutation and variation. *Am J Hum Genet* 91:1033–1040.
16. Polak P, R Karlić, A Koren, R Thurman, R Sandstrom, MS Lawrence, A Reynolds, E Rynes, K Vlahoviček, JA Stamatoyannopoulos and SR Sunyaev. (2015). Cell-of-origin chromatin organization shapes the mutational landscape of cancer. *Nature* 518:360–364.
 17. Pope BD, T Ryba, V Dileep, F Yue, W Wu, O Denas, DL Vera, Y Wang, RS Hansen, et al. (2014). Topologically associating domains are stable units of replication-timing regulation. *Nature* 515:402–405.
 18. Savage JRK. (2000). Proximity matters. *Science* 290:62–63.
 19. Misteli T. (2010). Higher-order genome organization in human disease. *Cold Spring Harb Perspect Biol* 2:a000794.
 20. Kinoshita T, G Nagamatsu, T Kosaka, K Takubo, A Hotta, J Ellis and T Suda. (2011). Ataxia-telangiectasia mutated (ATM) deficiency decreases reprogramming efficiency and leads to genomic instability in iPSC cells. *Biochem Biophys Res Commun* 407:321–326.
 21. Nayler S, M Gatei, S Kozlov, R Gatti, JC Mar, CA Wells, M Lavin and E Wolvetang. (2012). Induced pluripotent stem cells from ataxia-telangiectasia recapitulate the cellular phenotype. *Stem Cells Transl Med* 1:523–535.
 22. Fukawatase Y, M Toyoda, K Okamura, K Nakamura, K Nakabayashi, S Takada, M Yamazaki-Inoue, A Masuda, M Nasu, et al. (2014). Ataxia telangiectasia derived iPSC cells show preserved x-ray sensitivity and decreased chromosomal instability. *Sci Rep* 4:5421.
 23. Lee P, NT Martin, K Nakamura, S Azghadi, M Amiri, U Ben-David, S Perlman, RA Gatti, H Hu and WE Lowry. (2013). SMRT compounds abrogate cellular phenotypes of ataxia telangiectasia in neural derivatives of patient-specific hiPSCs. *Nat Commun* 4:1824.
 24. Lin L, MR Swerdel, MP Lazaropoulos, GS Hoffman, AJ Toro-Ramos, J Wright, H Lederman, J Chen, JC Moore and RP Hart. (2015). Spontaneous ATM gene reversion in A-T iPSC to produce an isogenic cell line. *Stem Cell Rep* 5:1097–1108.
 25. Park I, P Lerou, R Zhao, H Huo and G Daley. (2008). Generation of human-induced pluripotent stem cells. *Nat Protoc* 3:1180–1186.
 26. Liu P, A Kaplan, B Yuan, JH Hanna, JR Lupski and O Reiner. (2014). Passage number is a major contributor to genomic structural variations in mouse iPSCs. *Stem Cells* 32:2657–2667.
 27. Kastan MB and DS Lim. (2000). The many substrates and functions of ATM. *Nat Rev Mol Cell Biol* 1:179–186.
 28. Weissbein U, N Benvenisty and U Ben-David. (2014). Quality control: genome maintenance in pluripotent stem cells. *J Cell Biol* 204:153–163.
 29. Zhang Y, RP McCord, Y-J Ho, BR Lajoie, DG Hildebrand, AC Simon, MS Becker, FW Alt and J Dekker. (2012). Spatial organization of the mouse genome and its role in recurrent chromosomal translocations. *Cell* 148:908–921.
 30. Dileep V, JC Rivera-Mulia, J Sima and DM Gilbert. (2015). Large-scale chromatin structure-function relationships during the cell cycle and development: Insights from replication timing. *Cold Spring Harb Symp Quant Biol* [Epub ahead of print]; DOI: 10.1101/sqb.2015.80.027284.
 31. Hiratani I, T Ryba, M Itoh, T Yokochi, M Schwaiger, CW Chang, Y Lyou, TM Townes, D Schübeler and DM Gilbert. (2008). Global reorganization of replication domains during embryonic stem cell differentiation. *PLoS Biol* 6:e245.
 32. Ryba T, I Hiratani, J Lu, M Itoh, M Kulik, J Zhang, TC Schulz, AJ Robins, S Dalton and DM Gilbert. (2010). Evolutionarily conserved replication timing profiles predict long-range chromatin interactions and distinguish closely related cell types. *Genome Res* 20:761–770.
 33. Hajihosseini M, L Iavachev and J Price. (1993). Evidence that retroviruses integrate into post-replication host DNA. *EMBO J* 12:4969–4974.
 34. Humphries EH, C Glover and ME Reichmann. (1981). Rous sarcoma virus infection of synchronized cells establishes provirus integration during S-phase DNA synthesis prior to cellular division. *Proc Natl Acad Sci U S A* 78:2601–2605.
 35. Yant SR, X Wu, Y Huang, B Garrison, SM Burgess and MA Kay. (2005). High-resolution genome-wide mapping of transposon integration in mammals. *Mol Cell Biol* 25:2085–2094.
 36. Ruiz S, AJ Lopez-Contreras, M Gabut, RM Marion, P Gutierrez-Martinez, S Bua, O Ramirez, I Olalde, S Rodrigo-Perez, et al. (2015). Limiting replication stress during somatic cell reprogramming reduces genomic instability in induced pluripotent stem cells. *Nat Commun* 6:8036.
 37. Rube CE, S Zhang, N Miebach, A Fricke and C Rube. (2011). Protecting the heritable genome: DNA damage response mechanisms in spermatogonial stem cells. *DNA Repair* 10:159–168.
 38. Goodarzi AA, AT Noon, D Deckbar, Y Ziv, Y Shiloh, M Löbrich and PA Jeggo. (2008). ATM signaling facilitates repair of DNA double-strand breaks associated with heterochromatin. *Mol Cell* 31:167–177.
 39. Supek F and B Lehner. (2015). Differential DNA mismatch repair underlies mutation rate variation across the human genome. *Nature* 521:81–84.

Address correspondence to:

Dr. Paul H. Lerou
Division of Neonatology
Department of Pediatrics
Massachusetts General Hospital
Founders Building, 526C
55 Fruit Street
Boston, MA 02114

E-mail: plerou@partners.org

Received for publication December 23, 2015

Accepted after revision March 2, 2016

Prepublished on Liebert Instant Online March 2, 2016