Stem Cell Reports
Article

ISSCR

OPEN ACCESS

Spontaneous *ATM* Gene Reversion in A-T iPSC to Produce an Isogenic Cell Line

Lucy Lin,[1] Mavis R. Swerdel,[1] Michael P. Lazaropoulos,[1] Gary S. Hoffman,[1] Alana J. Toro-Ramos,[1] Jennifer Wright,[2] Howard Lederman,[2] Jianmin Chen,[1] Jennifer C. Moore,[1,3,4] and Ronald P. Hart[1,4,*]

[1]Department of Cell Biology and Neuroscience, Rutgers University, Piscataway, NJ 08854, USA
[2]A-T Clinic, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA
[3]Department of Genetics, Rutgers University, Piscataway, NJ 08854, USA
[4]Human Genetics Institute of New Jersey, Piscataway, NJ 08854, USA
*Correspondence: rhart@rutgers.edu
http://dx.doi.org/10.1016/j.stemcr.2015.10.010

SUMMARY

A spontaneously reverted iPSC line was identified from an A-T subject with heterozygous ATM truncation mutations. The reverted iPSC line expressed ATM protein and was capable of radiation-induced phosphorylation of CHK2 and H2A.X. Genome-wide SNP analysis confirmed a match to source T cells and also to a distinct, non-reverted iPSC line from the same subject. Rearranged T cell receptor sequences predict that the iPSC culture originated as several independently reprogrammed cells that resolved into a single major clone, suggesting that gene correction likely occurred early in the reprogramming process. Gene expression analysis comparing $ATM^{-/-}$ iPSC lines to unrelated $ATM^{+/-}$ cells identifies a large number of differences, but comparing only the isogenic pair of A-T iPSC lines reveals that the primary pathway affected by loss of ATM is a diminished expression of *p53*-related mRNAs. Gene reversion in culture, although likely a rare event, provided a novel, reverted cell line for studying ATM function.

INTRODUCTION

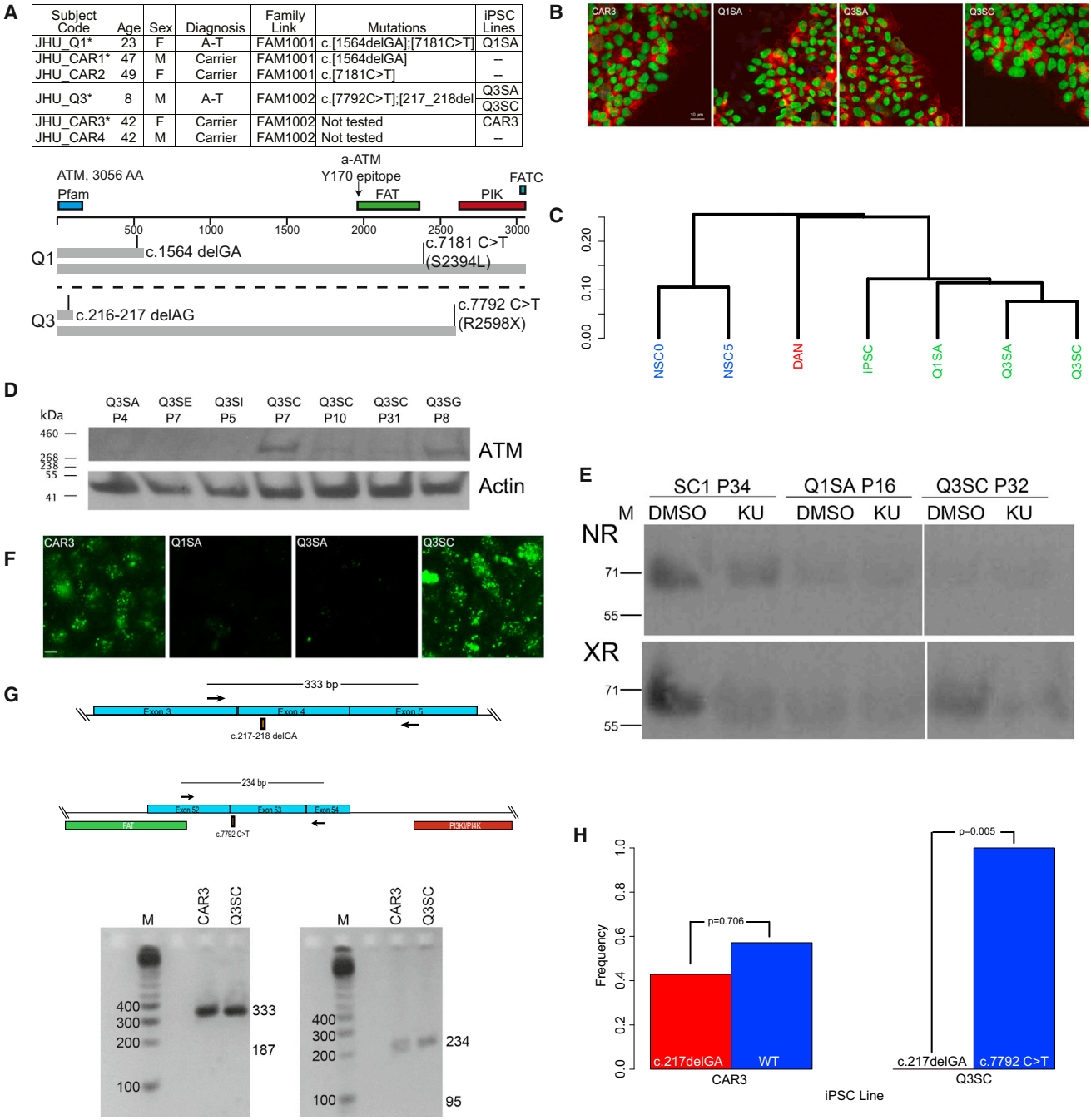
Ataxia-telangiectasia (A-T) is caused by the loss of function of the *ATM* gene, usually as a heterozygotic combination of two allelic mutations. In most cases, *ATM* mutations lead to expression of truncated proteins through early termination, but there are a large number of variations in mutation sites (Concannon and Gatti, 1997; Jacquemin et al., 2012). Studies of functional properties of A-T missense mutations found varying ATM protein levels and abnormal enrichment of cytoplasmic ATM protein (Jacquemin et al., 2012). Mouse models of A-T have been limited to standard knockout strategies that are normally homozygotic (Barlow et al., 1996; Xu et al., 1996). Because mutations may lead to expression of a portion of the ATM protein or even a full-length ATM protein lacking key regulatory sites, there is a possibility of partial function or dominant-negative function in A-T. For this reason, we sought to develop a collection of human A-T induced pluripotent stem cells (iPSC) for functional studies.

We previously discovered that ATM, in addition to serving many roles in cell-cycle regulation and DNA damage response, also functions as a key epigenetic regulator in neurons, likely underlying the ataxic symptoms as well as neurodegeneration. ATM indirectly regulates nuclear-cytoplasmic shuttling of HDAC4 (Li et al., 2012), and it directly phosphorylates EZH2, the enzymatic component of the PRC2 complex (Li et al., 2013). Each of these mechanisms contributes to ataxia symptoms in knockout mouse models, because reversal restores behavioral function. ATM also phosphorylates TET1, leading to conversion of 5mC to 5hmC marks, thought to be a key step in DNA demethylation (Jiang et al., 2015). Others have reported an effect on the DNMT1 methyltransferase via ATM phosphorylation of the Rb protein (Shamma et al., 2013). Any or all of these mechanisms could be affected or modulated by domains within a truncated ATM protein.

Although others have reported A-T iPSC (Nayler et al., 2012; Lee et al., 2013; Fukawatase et al., 2014), there has been no systematic preparation of a range of *ATM* mutation sites and no broad assessment of genome stability in iPSC cultures. It is expected that mutations occur spontaneously in tissue culture and presumably even in vivo. We examined several A-T iPSC lines and sublines derived from individually picked colonies for evidence of such reversion. We found that iPSC made from one subject yielded phenotypic and genotypic differences among sublines. Analysis of the resulting isogenic pair of cell lines greatly focuses the identification of cellular mechanisms, such as gene expression differences with or without ATM.

RESULTS

Blood samples were obtained from six subjects harboring previously identified *ATM* mutations (Figure 1A): two from subjects diagnosed with A-T (labeled as "Q") and four from carrier parents (labeled as "CAR"). iPSC were prepared from four of these subjects (asterisks in Figure 1A) using enriched, activated T cells and non-integrating Sendai



Subject Code	Age	Sex	Diagnosis	Family Link	Mutations	iPSC Lines
JHU_Q1*	23	F	A-T	FAM1001	c.[1564delGA];[7181C>T]	Q1SA
JHU_CAR1*	47	M	Carrier	FAM1001	c.[1564delGA]	--
JHU_CAR2	49	F	Carrier	FAM1001	c.[7181C>T]	--
JHU_Q3*	8	M	A-T	FAM1002	c.[7792C>T];[217_218del	Q3SA
						Q3SC
JHU_CAR3*	42	F	Carrier	FAM1002	Not tested	CAR3
JHU_CAR4	42	M	Carrier	FAM1002	Not tested	--

Figure 1. A-T iPSC Characterization

(A) Subject and iPSC line information. Six subjects were collected from the Johns Hopkins University Medical Center A-T Clinic with the characteristics shown. The names of iPSC lines derived from each subject are listed separately from source subject codes. *Subject samples used for iPSC reprogramming. The predicted patterns of ATM translation from each allele are shown for JHU_Q1 and JHU_Q3.

(B) Pluripotency markers in iPSC cultures. Example immunocytochemical images for each iPSC line showing nuclear OCT4 (green) and cell surface TRA-1-60 (red) staining. The scale bar represents 10 μm.

(C) Gene expression analysis (RNA-seq) results were clustered to show expression patterns consistent with pluripotency. Samples of unrelated iPSC, hESC-derived NSC at 0 or 5 days of neuronal differentiation (NSC0 and NSC5), and iPSC-derived midbrain-like DAN were included for comparison. NSC0, NSC5, DAN, and iPSC, n = 3; Q1SA, Q3SA, and Q3SC, n = 2.

(legend continued on next page)

viral vectors to deliver reprogramming factors (Moore et al., 2012). Sample iPSC colonies from each subject had standard morphology and stained positive for Oct4 and TRA-1-60, as shown in Figure 1B. For each subject, several sublines were picked from single colonies, expanded, and stored as frozen stocks. A subset of these was tested for gene expression patterns consistent with pluripotency. As shown in Figure 1C, three iPSC sublines (Q1SA, Q3SA, and Q3SC) all clustered with unrelated iPSC prepared from a non-A-T subject ("iPSC"), but they clustered separately from H1 human embryonic stem cell (hESC)-derived neural stem cells (NSC) at day 0 (NSC0) or NSC following 5 days of differentiation (NSC5) (Sauvageau et al., 2013), as well as dopaminergic neurons (DAN) differentiated from iPSC using the dual-SMAD protocol (Kriks et al., 2011). Furthermore, analysis of gene expression patterns using the PluriTest algorithm (Müller et al., 2011) identifies each iPSC line as being consistent with pluripotency (pluri-raw values ≥ 27, novelty ≤ 1.8). These results demonstrate successful iPSC reprogramming.

To confirm the presence of *ATM* gene mutations, we assessed ATM protein expression by western blot (Figures 1D, S1A, and S1B). A band corresponding to full-length ATM protein was clearly visible in the CAR3 and Q1SA iPSC lines, as expected (Figure S1A). Q1SA encodes a missense mutation (c.7181 C > T; encoding S2394L, Figure 1A) that is predicted to affect ATM activity but not translation of full-length protein. However, independent sublines prepared from subject JHU_Q3 expressed unexpected variation in ATM expression (Figures 1D and S1B). Some sublines had no ATM protein, as expected (sublines Q3SA, Q3SE, and Q3SI), but others expressed full-length protein (Q3SC and Q3SG). Furthermore, different passage numbers from the same subline exhibited varying quantities of ATM protein, peaking at P7 and then diminishing but not disappearing in later passages (Q3SC; Figures 1D and S1B). We selected two sublines, Q3SA and Q3SC, as contrasting examples without or with ATM expression, respectively. ATM activity in sample iPSC lines was confirmed by the X-irradiation (XR)-induced phosphorylation of CHK2 as detected by western blots (Figure 1E). The pCHK2 band was enhanced by XR (8 Gy) and diminished by pre-treatment with the ATM-specific inhibitor KU-55933 (KU) in unrelated control iPSC (SC1) and in Q3SC. Little pCHK2 was detected in Q1SA. Results were confirmed by observing the presence of XR-induced γH2A.X nuclear puncta with the same cell lines exhibiting positive (CAR3 and Q3SC) or negative (Q1SA and Q3SA) results (Figures 1F and S1C). Therefore, Q1SA expressed full-length ATM protein but had little or no kinase activity, presumably because of the presence of a previously unreported C > T variant at position c.7181, producing a S2394L that apparently affects activation. However, results indicate both the presence of ATM protein and its function in the Q3SC subline where none was expected on the basis of the diagnostic genotype provided with the subject cells. Results from Q3SA matched those expectations, as no ATM protein or activity was observed, with one allele terminating shortly after a frameshift mutation (c.217_218 delGA) and the other allele terminating at a point mutation (c.7792 C > T; R2598X), each truncating translation prior to the kinase domain.

Several possibilities could explain the aberrant expression of ATM in Q3SC. Exon skipping has been observed in several genes (Dutertre et al., 2010; Nevo et al., 2012) and is believed to be influenced by ATM function (Tresini et al., 2015). By amplifying cDNA with PCR primers targeting sites outside of exons which contain mutations in Q3 (sequences found in Table S1), only full-length product was identified, providing no evidence for mRNA skipping those exons (Figure 1G). We also counted the frequencies of sequenced, cloned PCR products corresponding to each allele and found only the c.7792 C > T allele mixed with wild-type sequence in Q3SC: there was no evidence of the c.217_218 delGA allele in the mRNA, as was found in CAR3 (Figure 1H). This would be consistent with gene reversion, because a simple deletion could not generate a wild-type allele.

The mutation sites were directly examined by sequencing of crude PCR-amplified segments of genome. As shown in Figure 2A, Q3SA exhibits the mutations identified in the diagnostic genotyping originally provided for

(D) Western blot to detect ATM protein expression in Q3-specific iPSC sublines. See also Figures S1A and S1B. Results were replicated in three western blots.

(E) Western blot to detect XR-induced pCHK2. Cultures were pre-treated with 10 μM KU or DMSO vehicle 1 hr prior to irradiation. For both NR and XR blots, the white space separates two regions from the same blot. Results were replicated in two western blots.

(F) Immunocytochemical detection of γH2A.X nuclear puncta induced by XR. Controls (NR and XR with/without KU inhibition) are shown in Figure S1C. The scale bar represents 10 μm.

(G) Assays to detect exon skipping in *ATM* mRNA. PCR primers were designed for exons upstream or downstream of mutations in JHU_Q3. The predicted product for skipping Exon 4 would be 187 bp and intact Exon 4 would be 333 bp. Only the latter size product was detected. Skipping Exon 53 would produce a 95 bp product, which was not observed.

(H) Allele frequencies in *ATM* mRNA. PCR products including the mutation sites were prepared from cDNA, cloned into plasmids to isolate individual molecules, and sequenced. The p value indicates the probability of observing these frequencies compared with an expected 50:50 ratio, according to chi-square analysis. The observed numbers of sequences were CAR3 c.216-217 delAG = 3, wild-type = 4; Q3SC c.216-217 delAG = 0, and c.7792 C > T = 8.

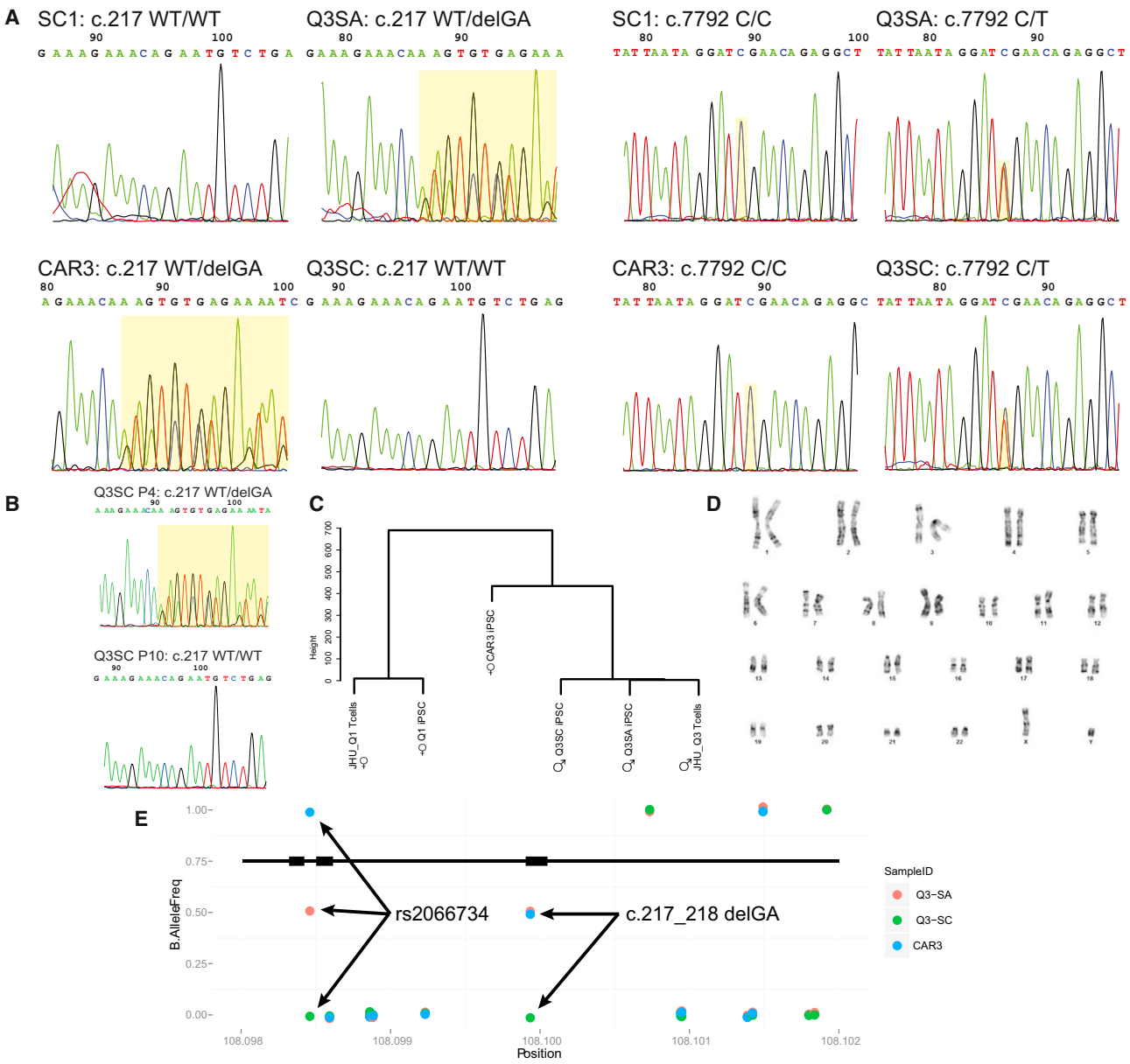


Figure 2. Genetic Analysis of A-T iPSC

(A) Sequencing traces of unpurified genomic PCR products. Homozygous sequences appear as a single pattern, and heterozygous sequences show a mixture of the two alleles distal from the point of difference (highlighted in yellow). The presence of the c.217_218 delGA deletion is visible as multiple overlapping traces in Q3SA and CAR3. The c.7792 C > T substitution is visible in Q3SA and Q3SC. Other samples were wild-type (similar to unrelated SC1) at these locations.

(B) Loss of the c.217_218 delGA variation over passaging of Q3SC. DNA samples from passage 4 (P4) or 10 (P10) identified the presence of this variation at P4 but not at P10.

(C) Hierarchical SNP clustering demonstrates genetic relationships among iPSC from genome-wide SNP arrays. Clustering used Euclidean distance by substituting 1 for AA, 0 for AB, and −1 for BB. Also indicated are the imputed genders based on SNPs.

(D) Karyotype of Q3SC. A representative chromosomal spread is shown for Q3SC, which was found to be 46 XY in 20 of 20 cells.

(E) Map of SNPs detected by re-sequencing genomic DNA near c.217_218 delGA. The positions of the deletion and the one SNP exhibiting LOH are identified by arrows.

subject JHU_Q3 (Figure 1A). Each mutated location was clearly heterozygous in Q3SA, as evidenced by the mixed trace from the unpurified PCR product (highlighted in yellow). CAR3 exhibited the c.217_218 delGA mutation, which is consistent with a parental relationship to JHU_Q3. Q3SC, on the other hand, is missing the c.217_218 delGA mutation, because only wild-type sequence is visible in the PCR products. The c.7792 C > T mutation is present in both Q3SA and Q3SC. The same pattern of mutations appears in sequenced, PCR-amplified segments of cDNA (not shown), suggesting that both alleles are transcribed. These results predict that Q3SC carries one allele capable of encoding a full-length ATM protein, while Q3SA does not.

Because the ATM western blot (Figure 1D) appears to show different quantities of ATM protein at different passage numbers of Q3SC, we wondered if early passage cells might have a different genotype from later passage cells. Early passage (P4) DNA prepared from Q3SC exhibited c.217_218 delGA in one allele, matching Q3SA and the diagnostic genotype for the JHU_Q3 subject, but later passage cells (P10) lacked this mutation (Figure 2B). This suggests that the Q3SC iPSC culture may have acquired a spontaneous gene reversion during early passaging. Alternatively, this culture could have suffered from cross-contamination. The latter seems unlikely, because the ATM mutation genotype for Q3SC does not match the maternal heterozygous carrier that we had thawed and reprogrammed (CAR3) but instead matches the imputed genotype and gender of cells from the father (CAR4), from whom we stored blood samples that have never been thawed or cultured in our laboratory.

To rule out the possibility of cross-contamination, we evaluated large numbers of genomic variations using SNP arrays. Clustering with the Euclidean distance between called SNPs, DNA from both Q3SA and Q3SC are quite similar to DNA prepared from the amplified JHU_Q3 T cells originally used for reprogramming (Q3SA: 699,368 SNPs match JHU_Q3 out of 701,230 called SNPs [99.73%]; Q3SC: 99.85%; Figure 2C). Less closely related is the DNA from the parent CAR3 (71.95%), the unrelated Q1SA iPSC line (55.65%), or its source JHU_Q1 T cells (55.64%). This confirms that Q3SA and Q3SC are iPSC lines made from the same source individual (JHU_Q3), who is related to JHU_CAR3, but not to the source T cells or iPSC prepared from JHU_Q1. Furthermore, on the basis of SNPs, Q3SA and Q3SC are confirmed as male and both Q1SA and CAR3 are confirmed as female, consistent with designated cell sources. SNPs overlapping those recommended for use in cell line identification (Yu et al., 2015) are shown in Table S2. The ATM-expressing Q3SC, then, is not due to cross-contamination of cultures.

We then searched for evidence of sequence variation in Q3SC. The karyotype of Q3SC was normal (Figure 2D; 46,XY in all 20 cells counted). SNP data were evaluated for evidence of loss of heterozygosity (LOH). Plots of the B allele frequency over the entire genome (not shown) or over the region of chr11 surrounding the *ATM* gene (Figure S2A) show no clear evidence for LOH overall but a lack of any heterozygous SNPs within the *ATM* gene, precluding the ability to assess LOH within the gene itself. However, the presence of SNP heterozygosity flanking *ATM* in all cell lines argues against a single crossing-over event between alleles, as this would have produced LOH distal from the event. By highlighting the 1,034 SNPs that distinguish Q3SA or Q3SC from source T cells (Figure S2B, colored dots) it is apparent that there is no concentrated genomic region of LOH between these two sublines, but that sequence variation has accumulated throughout the genome. Similar plots of the signal intensity of SNPs (log R ratio, not shown) show little evidence for copy-number variation (CNV) near the *ATM* gene. Not all known SNPs are included on the SNP array, so a 3.5 kb region surrounding the c.217_218 delGA site was re-sequenced. Only one SNP, rs2066734, was found to be heterozygous in Q3SA (red dot, Figure 2E). This SNP was homozygous for the major allele in Q3SC (green) and homozygous for the minor allele in CAR3 (blue). By estimating the rate of random SNP LOH using the SNP array data, a Fisher's exact test indicates that the probability of finding a specific LOH for rs2066734 within ATM but unrelated to the loss of c.217_218 delGA is $p = 0.00014$. This indicates that the most likely source of reversion was a short gene correction event, including at least ~1.5 kb upstream of c.217_218 delGA to include rs2066734. This result also confirms that the maternal allele, containing the c.217_218 delGA deletion as well as the minor rs2066734 allele, is lost in Q3SC and is replaced by a copy of the paternal allele sequence, explaining the absence of mutation at this position. Because mutation at c.7792 is heterozygous, reversion of c.217_218 delGA to wild-type sequence would be sufficient to produce a genotype of $ATM^{+/-}$.

To evaluate possible genomic rearrangement and CNV of *ATM* in individual cells, we used a fluorescence in situ hybridization (FISH) probe specific for the *ATM* gene (chr11:108,089,374–108,241,929 in the hg19 genome). Results indicated that a majority of cells had an equal number of *ATM*-specific puncta as those specific for a chr11 centromere control FISH probe (Figure 3A, "Normal"). A much smaller percentage of cells had one *ATM* singlet punctum and a doublet puncta compared with two singlets for chr11 control ("SD"), indicating asynchronous DNA replication. ATM-deficient cells have difficulties responding to oxidative stress in cell culture, and they may have diminished control of asynchronous DNA replication due

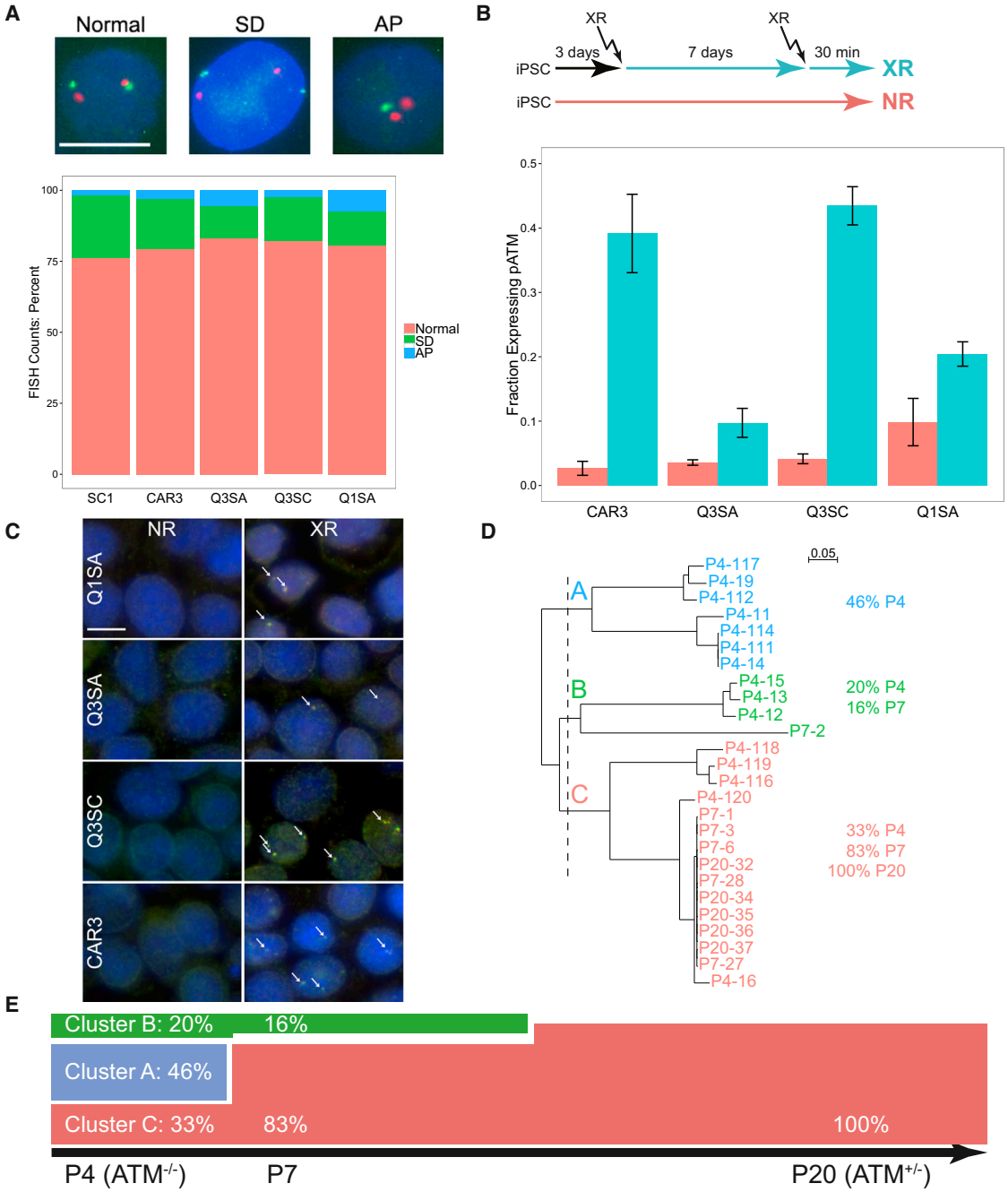


Figure 3. Genetic Variation in A-T iPSC

(A) FISH analysis identifies similar proportions of cells with normal, "SD" (unequal number of puncta for *ATM* compared with chr11 centromere), and "AP" (aneuploid) in CAR3, Q1SA, Q3SA, Q3SC, or unrelated control (SC1). Example images are shown for comparison (green, *ATM*; red, chr11 centromere). The scale bar represents 10 μm. Cells classified as "AP" are likely apoptotic. Total cells counted: SC1, n = 111; CAR3, n = 131; Q3SA, n = 123; Q3SC, n = 132; and Q1SA, n = 100.

(B) XR-induced pATM in iPSC following mutagenesis. Cultures were treated with 1 Gy XR once to induce mutagenesis and then again 1 week later to activate ATM autophosphorylation. The fraction of XR-induced pATM-positive nuclei is plotted for each iPSC line, with XR or without (NR) stimulation (n = 4 wells). This assay was repeated with cells from a different passage number with essentially identical results (not shown).

(C) Examples of immunocytochemically detected pATM and γH2A.X puncta are identified with arrows. The scale bar represents 10 μm.

(legend continued on next page)

to deficiencies in p53 signaling (Nagler et al., 2010). Rarely, cells were characterized as aneuploid ("AP") that likely reflected early stages of cell death. There were no significant differences in these percentages across genotype ($ATM^{-/-}$ versus $ATM^{+/-}$) or iPSC line. This indicates that no deletion or duplication of the *ATM* gene was detected, even in a subset of cells and therefore there was no evidence for cell-to-cell variability in *ATM* gene copy number.

Because the results suggested that gene correction may have occurred during the early passaging of Q3SC, we asked whether any of the $ATM^{-/-}$ iPSCs could be reverted in culture. To increase the chances of observing rare events, 96-well plates containing wells with CAR3, Q1SA, Q3SA, or Q3SC iPSC were X-irradiated at 1 Gy to promote DNA damage and induce genetic recombination or repair events. One week later, the cultures were irradiated again to activate ATM, and after 30 min of response time, all cultures were fixed, stained for phospho-ATM (pATM) and γH2A.X, and then imaged (Figure 3B). A parallel plate with no radiation treatment served as negative control. Automated counting of the fraction of pATM-containing nuclei indicated XR-induced ATM autophosphorylation. The $ATM^{+/-}$ lines, CAR3 and Q3SC, each had robust increases in pATM following XR. Q1SA, which encodes a full-length ATM protein with a S2394L variation, appeared to induce phosphorylation at S1981, the epitope for the phospho-specific ATM antibody, but this increase was similar to non-irradiated (NR) background. Similarly, the $ATM^{-/-}$ Q3SA appeared to show a weaker trend of XR-induced pATM activity that was not different from NR. To determine if any evidence for functional ATM could be detected, we searched images for coincident pATM and γH2A.X staining (Figure 3C). Q3SC and CAR3 had robust, overlapping puncta, as expected. However, rare nuclei could be identified in both Q1SA and Q3SA that had not only pATM puncta but also overlapping γH2A.X puncta, indicating that ATM was expressed, autophosphorylated, and likely functional in triggering H2A.X phosphorylation. In the case of Q1SA, it is possible that the variation at S2394 was not sufficient to completely inhibit ATM activation in all cells, even though assay of γH2A.X in the absence of XR was negative (Figure 1F). However, the presence of a weak signal of pCHK2 (Figure 1E) along with the trend of occasional pATM and γH2A.X activation in Figure 3C suggests that the variant of ATM expressed in Q1SA is capable of phosphorylating downstream targets in a small number of cells. However, detection of pATM and γH2A.X staining in a very small number of Q3SA cells could not be explained by the genotype of these cells, and therefore it is possible that the same type of gene reversion had occurred as that detected in Q3SC. These results indicate that the A-T iPSC are generally stable, at least over the period of this assay, but that XR-induced reversion can be detected at low frequencies in both Q1SA and Q3SA cells.

Because reversion is a very rare event we wondered how Q3SC transitioned so quickly from $ATM^{-/-}$ to $ATM^{+/-}$ (most likely between P4 and P10; see Figure 2B). Because these iPSC lines were prepared from T cells, we had a unique opportunity to trace individual subject cells during the reprogramming and growth stages in preparing iPSC. Mature T cells have rearranged genomic T cell receptor (*TCR*) and this rearrangement is unique to individual T cells. Therefore, we PCR-amplified genomic DNA from iPSC from various cell lines or passage numbers to detect rearranged *TCRβ* variable region using primers validated previously (Assaf et al., 2000). The PCR products were inserted into plasmids, and individual clones were sequenced. Alignment of these sequences from several passages of Q3SC indicates evidence for a varying number of "founder" T cell signatures in early passage that transitioned to a single type of signature in later passages (Figures 3D and S3A). This indicates that each iPSC line (derived from a single colony) was not produced from a single cell but instead included multiple "founder" cells, each presumably derived from independent reprogramming events. The sequences specific for Q3SC are distinct from those from Q3SA or Q1SA (Figure S3B), indicating that separate lines came from independent mixtures of founders. Q3SC results are summarized by the percentage of each of three major clusters by frequency of aligned *TCRβ* sequence (Figure 3E). Clusters A and B, present as the majority of detected sequences at P4, disappeared over further passaging. Cluster C expanded from 33% at P4 to 83% at P10 and to 100% at P20. This means that the ATM reversion event could have occurred either in the original T cells or early after reprogramming of iPSC in an individual "clone" of founder cells, which then expanded to overcome the others. Sequencing traces of crude, PCR-amplified DNA from P4 (Figure 2B) had no evidence for reverted c.217_218 delGA. If 33% of the DNA were indeed previously reverted, the mutated allele at c.217_218 delGA should constitute only 33% of the P4 mixed traces in Figure 2B instead of the 50% expected if the culture were entirely $ATM^{-/-}$, and this is nearly impossible to judge from a non-quantitative sequence trace. The most parsimonious interpretation is that reversion occurred early after

(D) Clustering of PCR-amplified *TCRβ* junction sequences from various passages of Q3SC. The vertical dashed line cleaves the dendrogram into three clusters, labeled A, B, and C. Sequence alignments matching this dendrogram are shown in Figure S3A. As a comparison, clustering samples from several iPSC lines show distinct patterns generally specific for each line (Figure S3B).
(E) Interpretation of cluster sequence frequencies calculated from results in (D).

reprogramming in the cluster C iPSC of Q3SC, possibly when this cluster was present as a single cell.

Because Q3SA and Q3SC represent an isogenic pair of iPSC lines from the same individual but with contrasting *ATM* genotypes ($ATM^{-/-}$ versus $ATM^{+/-}$, respectively), we examined gene expression differences using RNA sequencing (RNA-seq), evaluating matches at the transcript level. Relatively few mRNAs were significantly different in the isogenic Q3 pair (206, false discovery rate [FDR] $\leq$ 5%), as contrasted with a much larger number of differences between Q1SA and Q3SC (1,349), even though both contrasts compared $ATM^{-/-}$ with $ATM^{+/-}$ (Figures 4A–4C). This indicates that observed gene expression differences due to variations in genetic backgrounds generally overwhelm any analysis of single-gene, variant-specific differences. Out of the 206 transcripts different in the isogenic pair of cell lines (Table S3), 113 (54.8%) also differed between Q1SA and Q3SC (Figure 4A). On the other hand, a large proportion (601 of 1,349 [44.6%]) of transcripts differing between Q1SA and Q3SC were in common with the Q1SA versus Q3SA comparison (Figure 4B), even though the latter compared two different $ATM^{-/-}$ cell lines. Nayler et al. (2012) identified 7,920 genes as significantly different in A-T iPSC compared with wild-type cells using the same selection criteria. This list translated to 7,402 unique RefSeq transcript identifiers, and only 50 of these overlapped with the 206 Q3SA versus Q3SC transcripts found here. Three hundred sixty-four of Nayler's selected transcripts overlapped with the broader Q1SA versus Q3SC list, again indicating that comparisons between isogenic pairs are distinct from those between cell lines with different genetic backgrounds.

Looking to functional characterization of the transcripts selected as diminished in Q3SA compared with Q3SC, the top enriched pathway was p53 signaling ($p = 1.63 \times 10^{-4}$; Figure 4D). This is reinforced by upstream analysis, in which *TP53* mRNA was reduced ~24-fold in the absence of ATM, and 18 of 19 genes predicted to be transcriptionally regulated by p53 were affected ($p = 3.46 \times 10^{-5}$). Other pathways enriched by the reduced transcripts included IGF-1 signaling, acute phase response signaling, and the Wnt/Ca^{+2} pathway (Figure 4D). Few pathways were enriched in the list of transcripts increased in Q3SA compared with Q3SC, and these had no clear biological interpretation (Figure 4D). Under normal conditions in the absence of ATM stimuli such as double-stranded breaks, the primary cellular effect is a reduction in p53 signaling.

DISCUSSION

Spontaneous genetic variation has been described in cultured cells almost from the beginning of genetic analysis in cell culture. In one example, published in 1968, Atkins and Gartler proved that the mutation rate in culture is the same whether selection is applied or not (Atkins and Gartler, 1968). However, there has been relatively little report of spontaneous reversion in iPSC created from genetic variant cells, except for cases in which naturally occurring reversion of cultured primary cells have been reprogrammed into iPSC (Tolar et al., 2014; Maclean et al., 2012). Careful studies demonstrate that much of the variation between iPSC lines may be attributed to genetic variation in source skin fibroblasts (Abyzov et al., 2012) and that skin cells accumulate mutations in situ (Martincorena et al., 2015). Our results show that cultured iPSC have the capacity to undergo genetic rearrangement, in this case a reversion from $ATM^{-/-}$ to $ATM^{+/-}$. All available evidence, including at least the 1.4 kb spanning the c.217_218 delGA position to the location of rs2066734, supports the conclusion that gene conversion of *ATM* occurred, replacing the maternal allele with a copy of the paternal allele to create an allele encoding full-length ATM protein capable of phosphorylating pCHK2 and γH2A.X upon XR. This likely occurred between initial reprogramming of the iPSC line (P0) and P10, as DNA prepared from P4 had not reverted, or at least not detectably. One puzzling aspect was that the ATM protein levels detected by western blotting appeared to change over passaging (Figure 1D). It is not clear if this is meaningful in the context of our study, as ATM protein remained detectable after the reversion event, even if levels were reduced. Alternatively, *ATM* gene reversion could have occurred in circulating T cells prior to reprogramming. Even if this were the case, Figure S2B demonstrates that independent SNP variants accumulate after reprogramming, proving that iPSC cultures accumulate genetic variants during culture.

Analysis of *TCRβ* signatures predicts that several initial "founder" clones of iPSC were present at P4 but that these resolved to a single clone by P20. This was somewhat surprising, as it is often assumed that individually picked colonies of iPSC are derived from a single cell and are therefore clonal. Our results demonstrate that this is not the case. In other studies, we found that multiple founder cells are found in most iPSC colonies using both T cell receptor rearrangement and by mixing source cells of different gender. There was no overt selection for ATM reversion among the multiple founder cell-derived iPSC, although we presume that the gain of ATM function, with its diverse roles regulating cell cycle, DNA repair, and epigenetic mechanisms, may have a selective advantage in the stressful environment of culture.

Many alternative explanations for this change in genotype were considered and discarded. No evidence for exon skipping or allele-specific mRNA expression could be detected. SNP analysis proves that the reverted cells

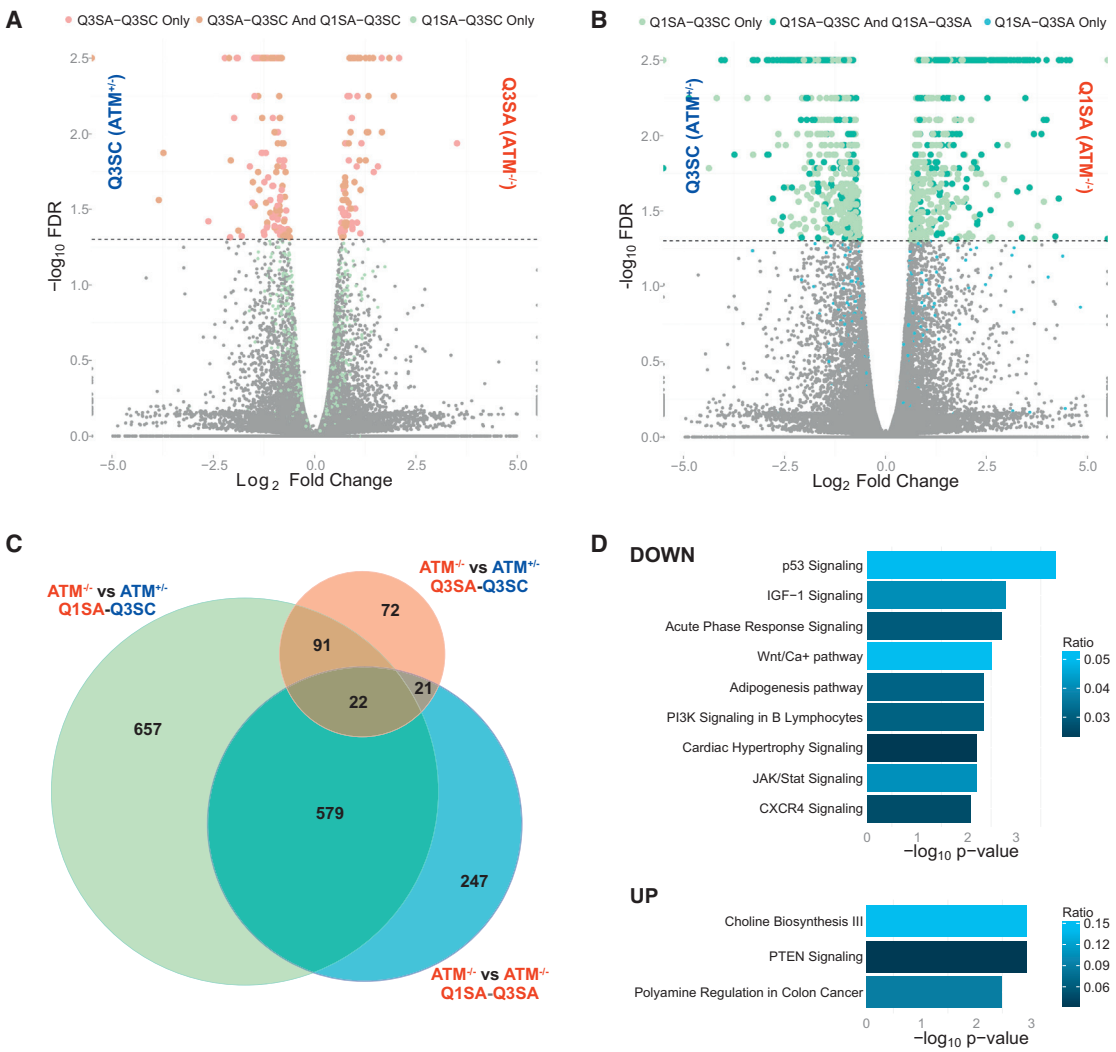


Figure 4. Gene Expression Analysis in A-T iPSC

(A) Volcano plot comparing changes in transcript (as $\log_2$ fold change of fragments per kilobase of transcript per million mapped reads) and significance ($-\log_{10}$ of the FDR) for Q3SA versus Q3SC. Sample names are color-coded by genotype: red for $ATM^{-/-}$ and blue for $ATM^{+/-}$. The significance threshold of 5% FDR is shown as a horizontal dashed line (n = 2 per group). Smaller gray dots indicate transcripts with no significant difference. Colors of dots indicate whether transcripts overlap Q1SA versus Q3SC or not: see key and the colors of overlapping regions in the Venn diagram (C) for color interpretation.

(B) Volcano plot comparing Q1SA with Q3SC, with dot coloring to show overlap with Q1SA versus Q3SC.

(C) Venn diagram indicating the numbers of significantly different transcripts in each contrast and the numbers of overlaps among the contrasts.

(D) Pathway analysis of transcripts that are decreased in abundance in Q3SA compared with Q3SC (top) or increased (bottom). The bar indicates the p value calculated by Fisher's exact test, and the color intensity depicts the ratio of pathway genes included in this differentially expressed set.

matched the source subject T cells as well as another iPSC subline that had not reverted. Other types of crossover events should have produced large regions of LOH or changes in copy number, and neither of these was found.

An attempt to induce reversion with XR-induced mutagenesis revealed isolated cases of cells staining positive for both pATM and γH2A.X in Q3SA cultures that should carry two mutated alleles to produce only truncated proteins. These small numbers of apparently reverted cells did not rise to the level of significance when assessed by automated counting of all available wells, so one might argue that these spontaneous events occur at an exceedingly low

level, if at all, in the absence of mutagenesis. On the other hand, we found pATM and γH2A.X staining cells in Q1SA, which expresses a missense mutation from one allele. These cells, harboring a variant encoding S2394L, could potentially have destroyed a site of autophosphorylation, but a previous search did not identify this serine residue as phosphorylated (Kozlov et al., 2011). On the basis of the possibility that this variant was "leaky" for ATM function, we could not distinguish if actual gene reversion had occurred or if ATM had been activated in only a small proportion of cells. The absence of γH2A.X staining without the mutagenic XR (Figure 1F) argues against this, but there have been cases reported of incomplete inactivation of ATM with other mutations (Stewart et al., 2001). The fact that both Q3SA and Q1SA cultures have been passaged in our lab for several months argues that if reversion does occur spontaneously, there is unlikely to be a strong functional selection for the reversion, because we have not yet observed a change in apparent genotype.

Having the unexpected but fortuitous case of two iPSC lines made from the same individual but with contrasting *ATM* genotypes, we assessed baseline gene expression patterns. The results clearly indicate that the mRNA patterns affected by the lack of ATM function primarily relate to p53 pathways. In support of this interpretation, upstream regulator analysis also identified a pattern of regulation not only of the gene encoding p53 (*TP53*) but also the majority of transcriptional targets downstream of this key protein. It will be invaluable to continue these analyses in future studies to identify patterns of mRNA changes following the induction of DNA damage. Most important, however, the regulated mRNA patterns comparing Q3SA with Q3SC are much more focused than a comparison between unrelated cultures having the same genotypes (Q1SA versus Q3SC). Previous studies (Nayler et al., 2012) analyzed similarly diverse A-T versus control cultures and concluded that other pathways, including mitochondrial function, might be directly affected by the loss of ATM function. Our results, although identifying a core set of genes in common with those of Nayler's study (Nayler et al., 2012), show that only by contrasting an isogenic pair of cells can the clearest functional patterns be revealed.

Reversion of an A-T iPSC line proved to be a valuable accident. However, the fact that an iPSC line derived from an A-T subject reverted at all provides a cautionary tale for those culturing gene variants in stem cells. Clearly, variations occur randomly in culture. As an example (Figure S2B), two different sublines prepared from the same individual exhibited 1,034 SNP variations from source cells independent of the *ATM* gene reversion. This rate of variation predicts that as many as 4×10^6 bp may vary between these lines across the entire genome. Clearly it is important that a gene variation of interest be monitored over growth of iPSC cultures, just as it is important to monitor cultures for evidence of cross-contamination (Yu et al., 2015). These results demonstrate that iPSC cultures must be evaluated continuously to ensure that the cells appropriately represent the genetic model being investigated.

EXPERIMENTAL PROCEDURES

Cell Source and Reprogramming

Standard blood samples were obtained from the Johns Hopkins A-T Clinic, with approval from the Johns Hopkins IRB (#NA_00077304), and shipped to RUCDR Infinite Biologics. Reprogramming from activated T cells using Sendai viral vectors expressing Oct4, Sox2, Klf4, and Myc (Cytotune; Life Technologies) followed protocols described previously (Moore et al., 2012). All iPSC cultures are grown under low (4%) O_2 conditions (Guo et al., 2013). Procedures at Rutgers were confirmed as exempt by the Rutgers institutional review board (#E13-825). Individual "sublines" were picked from a single, primary colony of iPSC.

RNA-seq

RNA was prepared from frozen iPSC pellets using Trizol (Life Technologies). RNA-seq was performed at RUCDR Infinite Biologics using Illumina technology and analyzed as described previously (Kraushar et al., 2014). The accession number for the A-T cell line RNA-seq experiments reported in this paper is GEO: GSE71878. The RNA samples used for comparison in Figure 1B were obtained from a previous study (NSC; Sauvageau et al., 2013) or from unpublished results from our lab.

Immunocytochemistry

Cultured cells were fixed in 4% paraformaldehyde. Primary antibodies were obtained from EMD Millipore (Oct-4 [clone 10H11.2], TRA-1-60 [MAB4360], γH2A.X [phospho-Ser139]) or from Abcam (P-ATMs1981 [ab36810]). Imaging was performed on an IN CELL 6000 automated confocal microscope system (GE Biosciences).

Western Blots

Proteins were extracted, run on precast NuPage 3%–8% acrylamide gels, transferred to Immun-Blot PVDF membrane (Bio-Rad), and detected with primary antibody and HRP-coupled secondary using Luminol reagent (Santa Cruz Biotechnology). Primary antibodies were obtained from Abcam (ATM [Y170, 2C1A1]), Cell Signaling Technology (Danvers MAT) (pCHK2 [phospho T68]), and Santa Cruz (actin [SC-1615]).

PCR and Sequencing

Genomic DNA was prepared using Zymo gDNA columns (Zymo Research), and specific regions were amplified by PCR using primers listed in Table S1 using Q5 mastermix (New England Biolabs). PCR fragments were cloned using the StrataClone Blunt PCR cloning kit (Agilent). Sequencing was performed by GenScript. ATM mutations were compared with annotations in the LOVD database (http://atm.lovd.nl). TCRβ junction sequences were identified by amplification with the Jβ1 family of primers

and the Vβ pan primer from Assaf et al. (2000). Amplified fragments were cloned into the Stratagene Blunt PCR cloning kit (Agilent) and sequenced with an M13R vector-specific primer.

SNP Arrays

Genomic DNA samples were run on Illumina Human OmniExpress SNP arrays at RUCDR Infinite Biologics. Data were analyzed using Illumina GenomeStudio software and R/Bioconductor.

Karyotyping

Karyotyping was performed by RUCDR Infinite Biologics.

FISH

Fluorescent SureFISH probes were purchased from Agilent Technologies (SureFISH 11q22.3 *ATM* 153kb and SureFISH Chr11 CEP 798kb). iPSC were spread on positively charged glass slides using protocol reported previously (Moralli et al., 2011). FISH was performed following Agilent protocol. Probes were visualized by a Leica DM5000B fluorescence microscope. Probe signals were enumerated in 200 cells for each iPSC subline.

SUPPLEMENTAL INFORMATION

Supplemental Information includes three figures and three tables and can be found with this article online at http://dx.doi.org/10.1016/j.stemcr.2015.10.010.

AUTHOR CONTRIBUTIONS

L.L., M.R.S., M.P.L., G.S.H., A.J.T.-R., J.C., and J.C.M. performed experiments. J.W. and H.L. selected subjects and collected A-T blood samples. R.P.H. designed the study, analyzed data, and wrote the manuscript.

ACKNOWLEDGMENTS

This work was supported by the A-T Children's Project. We thank Drs. Lourdes Serrano and Berta Nieves Vazquez Prat for sharing their optimized ATM western blotting methods, Dr. Steven Buyske for help with statistics, and Dr. Karl Herrup for many helpful discussions.

Received: September 18, 2015
Revised: October 16, 2015
Accepted: October 19, 2015
Published: November 19, 2015

REFERENCES

Abyzov, A., Mariani, J., Palejev, D., Zhang, Y., Haney, M.S., Tomasini, L., Ferrandino, A.F., Rosenberg Belmaker, L.A., Szekely, A., Wilson, M., et al. (2012). Somatic copy number mosaicism in human skin revealed by induced pluripotent stem cells. Nature *492*, 438–442.

Assaf, C., Hummel, M., Dippel, E., Goerdt, S., Müller, H.H., Anagnostopoulos, I., Orfanos, C.E., and Stein, H. (2000). High detection rate of T-cell receptor beta chain rearrangements in T-cell lymphoproliferations by family specific polymerase chain reaction in combination with the GeneScan technique and DNA sequencing. Blood *96*, 640–646.

Atkins, J.H., and Gartler, S.M. (1968). Development of a nonselective technique for studying 2, 6, diaminopurine resistance in an established murine cell line. Genetics *60*, 781–792.

Barlow, C., Hirotsune, S., Paylor, R., Liyanage, M., Eckhaus, M., Collins, F., Shiloh, Y., Crawley, J.N., Ried, T., Tagle, D., and Wynshaw-Boris, A. (1996). Atm-deficient mice: a paradigm of ataxia telangiectasia. Cell *86*, 159–171.

Concannon, P., and Gatti, R.A. (1997). Diversity of ATM gene mutations detected in patients with ataxia-telangiectasia. Hum. Mutat. *10*, 100–107.

Dutertre, M., Sanchez, G., De Cian, M.C., Barbier, J., Dardenne, E., Gratadou, L., Dujardin, G., Le Jossic-Corcos, C., Corcos, L., and Auboeuf, D. (2010). Cotranscriptional exon skipping in the genotoxic stress response. Nat. Struct. Mol. Biol. *17*, 1358–1366.

Fukawatase, Y., Toyoda, M., Okamura, K., Nakamura, K., Nakabayashi, K., Takada, S., Yamazaki-Inoue, M., Masuda, A., Nasu, M., Hata, K., et al. (2014). Ataxia telangiectasia derived iPS cells show preserved x-ray sensitivity and decreased chromosomal instability. Sci. Rep. *4*, 5421.

Guo, C.W., Kawakatsu, M., Idemitsu, M., Urata, Y., Goto, S., Ono, Y., Hamano, K., and Li, T.S. (2013). Culture under low physiological oxygen conditions improves the stemness and quality of induced pluripotent stem cells. J. Cell. Physiol. *228*, 2159–2166.

Jacquemin, V., Rieunier, G., Jacob, S., Bellanger, D., d'Enghien, C.D., Laugé, A., Stoppa-Lyonnet, D., and Stern, M.H. (2012). Underexpression and abnormal localization of ATM products in ataxia telangiectasia patients bearing ATM missense mutations. Eur. J. Hum. Genet. *20*, 305–312.

Jiang, D., Zhang, Y., Hart, R.P., Chen, J., Herrup, K., and Li, J. (2015). Alteration in 5-hydroxymethycytosine-mediated epigenetic regulation leads to Purkinje cell vulnerability in ATM deficiency. Brain http://dx.doi.org/10.1093/brain/awv284.

Kozlov, S.V., Graham, M.E., Jakob, B., Tobias, F., Kijas, A.W., Tanuji, M., Chen, P., Robinson, P.J., Taucher-Scholz, G., Suzuki, K., et al. (2011). Autophosphorylation and ATM activation: additional sites add to the complexity. J. Biol. Chem. *286*, 9107–9119.

Kraushar, M.L., Thompson, K., Wijeratne, H.R., Viljetic, B., Sakers, K., Marson, J.W., Kontoyiannis, D.L., Buyske, S., Hart, R.P., and Rasin, M.R. (2014). Temporally defined neocortical translation and polysome assembly are determined by the RNA-binding protein Hu antigen R. Proc. Natl. Acad. Sci. U S A *111*, E3815–E3824.

Kriks, S., Shim, J.W., Piao, J., Ganat, Y.M., Wakeman, D.R., Xie, Z., Carrillo-Reid, L., Auyeung, G., Antonacci, C., Buch, A., et al. (2011). Dopamine neurons derived from human ES cells efficiently engraft in animal models of Parkinson's disease. Nature *480*, 547–551.

Lee, P., Martin, N.T., Nakamura, K., Azghadi, S., Amiri, M., Ben-David, U., Perlman, S., Gatti, R.A., Hu, H., and Lowry, W.E. (2013). SMRT compounds abrogate cellular phenotypes of ataxia telangiectasia in neural derivatives of patient-specific hiPSCs. Nat. Commun. *4*, 1824.

Li, J., Chen, J., Ricupero, C.L., Hart, R.P., Schwartz, M.S., Kusnecov, A., and Herrup, K. (2012). Nuclear accumulation of HDAC4 in ATM

deficiency promotes neurodegeneration in ataxia telangiectasia. Nat. Med. *18*, 783–790.

Li, J., Hart, R.P., Mallimo, E.M., Swerdel, M.R., Kusnecov, A.W., and Herrup, K. (2013). EZH2-mediated H3K27 trimethylation mediates neurodegeneration in ataxia-telangiectasia. Nat. Neurosci. *16*, 1745–1753.

Maclean, G.A., Menne, T.F., Guo, G., Sanchez, D.J., Park, I.H., Daley, G.Q., and Orkin, S.H. (2012). Altered hematopoiesis in trisomy 21 as revealed through in vitro differentiation of isogenic human pluripotent cells. Proc. Natl. Acad. Sci. U S A *109*, 17567–17572.

Martincorena, I., Roshan, A., Gerstung, M., Ellis, P., Van Loo, P., McLaren, S., Wedge, D.C., Fullam, A., Alexandrov, L.B., Tubio, J.M., et al. (2015). Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin. Science *348*, 880–886.

Moore, J.C., Sheldon, M.H., and Hart, R.P. (2012). Biobanking in the Era of the Stem Cell: A Technical and Operational Guide (Morgan & Claypool Life Sciences).

Moralli, D., Yusuf, M., Mandegar, M.A., Khoja, S., Monaco, Z.L., and Volpi, E.V. (2011). An improved technique for chromosomal analysis of human ES and iPS cells. Stem Cell Rev. *7*, 471–477.

Müller, F.J., Schuldt, B.M., Williams, R., Mason, D., Altun, G., Papapetrou, E.P., Danner, S., Goldmann, J.E., Herbst, A., Schmidt, N.O., et al. (2011). A bioinformatic assay for pluripotency in human cells. Nat. Methods *8*, 315–317.

Nagler, A., Cytron, S., Mashevich, M., Korenstein-Ilan, A., and Avivi, L. (2010). The aberrant asynchronous replication—characterizing lymphocytes of cancer patients—is erased following stem cell transplantation. BMC Cancer *10*, 230.

Nayler, S., Gatei, M., Kozlov, S., Gatti, R., Mar, J.C., Wells, C.A., Lavin, M., and Wolvetang, E. (2012). Induced pluripotent stem cells from ataxia-telangiectasia recapitulate the cellular phenotype. Stem Cells Transl. Med. *1*, 523–535.

Nevo, Y., Kamhi, E., Jacob-Hirsch, J., Amariglio, N., Rechavi, G., Sperling, J., and Sperling, R. (2012). Genome-wide activation of latent donor splice sites in stress and disease. Nucleic Acids Res. *40*, 10980–10994.

Sauvageau, M., Goff, L.A., Lodato, S., Bonev, B., Groff, A.F., Gerhardinger, C., Sanchez-Gomez, D.B., Hacisuleyman, E., Li, E., Spence, M., et al. (2013). Multiple knockout mouse models reveal lincRNAs are required for life and brain development. eLife *2*, e01749.

Shamma, A., Suzuki, M., Hayashi, N., Kobayashi, M., Sasaki, N., Nishiuchi, T., Doki, Y., Okamoto, T., Kohno, S., Muranaka, H., et al. (2013). ATM mediates pRB function to control DNMT1 protein stability and DNA methylation. Mol. Cell. Biol. *33*, 3113–3124.

Stewart, G.S., Last, J.I., Stankovic, T., Haites, N., Kidd, A.M., Byrd, P.J., and Taylor, A.M. (2001). Residual ataxia telangiectasia mutated protein function in cells from ataxia telangiectasia patients, with 5762ins137 and 7271T–>G mutations, showing a less severe phenotype. J. Biol. Chem. *276*, 30133–30141.

Tolar, J., McGrath, J.A., Xia, L., Riddle, M.J., Lees, C.J., Eide, C., Keene, D.R., Liu, L., Osborn, M.J., Lund, T.C., et al. (2014). Patient-specific naturally gene-reverted induced pluripotent stem cells in recessive dystrophic epidermolysis bullosa. J. Invest. Dermatol. *134*, 1246–1254.

Tresini, M., Warmerdam, D.O., Kolovos, P., Snijder, L., Vrouwe, M.G., Demmers, J.A., van IJcken, W.F., Grosveld, F.G., Medema, R.H., Hoeijmakers, J.H., et al. (2015). The core spliceosome as target and effector of non-canonical ATM signalling. Nature *523*, 53–58.

Xu, Y., Ashley, T., Brainerd, E.E., Bronson, R.T., Meyn, M.S., and Baltimore, D. (1996). Targeted disruption of ATM leads to growth retardation, chromosomal fragmentation during meiosis, immune defects, and thymic lymphoma. Genes Dev. *10*, 2411–2422.

Yu, M., Selvaraj, S.K., Liang-Chu, M.M., Aghajani, S., Busse, M., Yuan, J., Lee, G., Peale, F., Klijn, C., Bourgon, R., et al. (2015). A resource for cell line authentication, annotation and quality control. Nature *520*, 307–311.