IJC
International Journal of Cancer

Major multilevel molecular divergence between THP-1 cells from different biorepositories

Nandita Noronha [1], Grégory Ehx [1], Marie-Christine Meunier [2], Jean-Philippe Laverdure [1], Catherine Thériault [1] and Claude Perreault [1,3]

[1]Institute for Research in Immunology and Cancer (IRIC), Université de Montréal, Montreal, QC, Canada
[2]HLA Laboratory, Hôpital Maisonneuve-Rosemont, Montreal, QC, Canada
[3]Department of Medicine, Université de Montréal, Montreal, QC, Canada

The THP-1 cell line is broadly used as a model for acute myeloid leukemia (AML) with *MLL* fusion and to study monocyte differentiation and function. We studied THP-1 cells obtained from two major biorepositories. The two cell lines were closely related with a percentage match of short tandem repeat (STR) profiles ranging from 93.75% to 100%, depending on the algorithm used. Nevertheless, we found that the two cell lines presented discordant HLA type, cytogenetic aberrations and AML-related gene expression (including critical targets of *MLL* fusion). These discrepancies resulted mainly from loss of heterozygosity (LOH) involving five chromosomal regions. In view of their aberrant expression of key "leukemia" genes (e.g., *LIN28B*, *MEIS1* and *SPARC*), we argue that one of the THP-1 cell lines may not be a reliable model for studying leukemia. Their defective expression of HLA molecules and abnormal adhesion properties is also a caveat for studies of antigen presentation. In a more general perspective, our findings show that seemingly minor discrepancies in STR profiles among cell lines may be the sign of major genetic drift, of sufficient magnitude to affect the reliability of cell line-based research.

Introduction

Acute myeloid leukemia (AML) is a malignant hematological disorder characterized by the proliferation of non-functional hematopoietic cells. More than 70% of infant AML cases bear a chromosomal translocation involving the *MLL* gene,[1] whose fusion product is instrumental in leukemogenesis.[2] Established in 1980 from the blood of a child with AML, THP-1 is one of the most widely used cell lines to study the biology of AML with *MLL* fusion as well as monocyte function.[3] Highly appreciated for its versatility, this cell line has been used in over 10,000 publications.

The scientific community is currently going through a well-documented "crisis" of reproducibility as more than 70% of researchers fail to reproduce another scientist's experiments, and more than half failed to reproduce their own.[4] It is well recognized that cross-contamination and phenotypic drift of cells in culture can generate irreproducible or misleading data.[5–9] Cell lines with microsatellite instability (MSI+), in particular, have been shown to develop alterations in microsatellite loci and drift with passages, due to their defective mismatch repair genes.[10] This biological diversity clearly contributes to the reproducibility crisis and is disquieting considering that human cancer cell lines are the workhorse of cancer research. The substantial biological diversity among genetically unstable cancer cell lines within and between labs has been extensively reported.[11,12] However, little is known about the possible diversity of cell lines between reference biorepositories. Here, we compared THP-1 cells, a cell line considered as relatively stable,[13] obtained from two different major biorepositories. We found that they were not the same entity and have undergone genetic drift with major and pervasive functional effects.

Materials and Methods

Cell lines

THP-1 cell lines (RRID: CVCL_0006) were freshly purchased from the American Type Culture Collection (ATCC; TIB-202, ATCC, Manassas, VA) and the Deutsche Sammlung von Mikroorganismen und Zellkulturen (DSMZ; ACC16, DSMZ, Braunschweig, Germany) for the present study. Cells were maintained in RPMI 1640 (Gibco, New York, NY, 11875-093) containing L-glutamine and supplemented with

N.N. and G.E. contributed equally to this work

Additional Supporting Information may be found in the online version of this article.

Key words: acute myeloid leukemia, DNA copy number variations, histocompatibility, reproducibility of results, THP-1 cells, transcriptomics

Abbreviations: AML: acute myeloid leukemia; ATCC: American Type Culture Collection; DEG: differentially expressed genes; DSMZ: Deutsche Sammlung von Mikroorganismen und Zellkulturen; FC: fold change; GSEA: gene set enrichment analysis; HTC: hotspots of transcriptomic changes; LOH: a loss of heterozygosity; STR: short tandem repeat

DOI: 10.1002/ijc.32967
History: Received 28 Nov 2019; Accepted 2 Mar 2020; Online 12 Mar 2019

Correspondence to: Claude Perreault,
E-mail: claude.perreault@umontreal.ca

What's new?

Human monocytic THP-1 cells are widely used for the study of myeloid leukemias. It remains unclear, however, to what degree THP-1 cell lines available from biorepositories differ. Here, THP-1 cells obtained from two major repositories were investigated using a multi-omics approach. The two cell lines were found to have numerous genomic, transcriptomic, and proteomic discrepancies, with pervasive effects on genes that serve critical roles in leukemogenesis. The findings indicate that the two THP-1 cell lines differ markedly, with evidence of biologically significant genetic drift that is likely to be underestimated by analyses of short tandem repeats.

20% heat-inactivated fetal bovine serum (FBS, Gibco 12483) and 1% penicillin–streptomycin (10,000 U/ml, Gibco 15140-122). All cultures were free of mycoplasma contamination, as verified by the absence of RNA-Seq reads (see in next sections) mapping to four different mycoplasma genomes (analysis made as reported previously[14]). Cell line authentication was performed by STR profiling. DNA was extracted from both cell lines and submitted for STR profiling to ATCC's cell authentication service (July 2019). STR profiling was performed based on 13 Combined DNA Index System (CODIS: D18S51, D21S11, TH01, D3S1358, FGA, TPOX, D8S1179, vWA, CSF1PO, D16S539, D7S820, D13S317, D5S818) loci plus amelogenin (for gender determination), Penta E, Penta D, D2S1338 and D19S433 (results reported in supplemental data and other data can be made available upon reasonable request).

HLA genotyping of THP-1 cell lines

Cells were genotyped at 6 HLA genes: HLA-A, HLA-B, HLA-C and HLA-DRB1, HLA-DQB1, HLA-DPB1. Typing was performed by next-generation sequencing using NGSgo® commercial kit (GenDx, Netherlands) on the MiSeq platform (Illumina, San Diego, CA). Paired-end sequences were analyzed using NGSengine® software v2.12.0 (GenDx) and the IPD-IMGT/HLA database release 3.34.0.

RNA sequencing

After thawing, THP-1 cells (ATCC and DSMZ) were expanded for 1 week to recover from cryopreservation. Three replicate expansions were performed for sample collection with the following sequences: seeding at 0.2 million/ml, 48 hr of expansion, dilution 1:2 with fresh medium, 24 hr of expansion and collection while in the log phase (final cell density = 0.5 million/ml). Replicates 1 and 3 were collected 1 week after thawing and Replicate 2 was collected 2 weeks after thawing.

Total RNA was extracted using TRIzol® (Life Technologies, Carlsbad, CA) according to the manufacturer's recommendations. RNA samples were purified using QIAGEN Mini RNeasy kit following manufacturer's instructions. The presence of contaminants was assessed with nanodrop using 260/280 and 260/230 ratios. RNA quantification was performed using QuBit (ABI) and 1 μg of total RNA was used for library preparation. Sample quality was assessed with Bioanalyzer Nano (Agilent Technologies, Santa Clara, CA) and all samples had RNA Integrity Numbers above 8.8. Transcriptomic libraries were prepared with the KAPA mRNA HyperPrep Kit (KAPA, Cat no. KR1352). Libraries were quantified with the QuBit and BioAnalyzer (average size ~380 base pairs). All libraries were diluted to 10 nM and normalized by quantitative PCR using the KAPA library quantification kit (KAPA; Cat no. KK4973). Libraries were pooled to equimolar concentrations and sequencing was performed with the Illumina Nextseq500 on half a flow cell of the Nextseq High Output v2 (75 cycles) kit using 4 nM of the pooled libraries. Around 40 M single-end reads passing filter were generated per sample. Library preparation and sequencing were performed at the genomics platform of the Institute for Research in Immunology and Cancer.

Transcriptomic analysis

Adapters were trimmed using trimmomatic 0.35 and transcript quantification was performed using Kallisto v0.43.0 with the '–single -l 300 -s 50 –rf-stranded' parameters (GRCh38.88). Analysis of differentially expressed genes (DEGs) was done in R3.5.1. Raw read counts have been converted to counts per million (cpm), normalized relative to the library size and lowly expressed genes were filtered out by keeping genes with cpm >1 in at least 2 samples using edgeR and limma. This was followed by voom transformations and linear modeling using limma's lmfit. Finally, moderated t-statistics were computed with eBayes. Genes with p values <0.05, fold change (FC) >2 and false discovery rate <0.01 were considered significantly differentially expressed. Gene ontology and biological pathways annotations were made with DAVIDv6.8 (https://david.ncifcrf.gov). Functional annotations with p-value < 0.05 were considered significant. Gene set enrichment analysis (GSEA) was performed with fgsea package in R.[15] A pre-ranked gene list was generated by ranking expressed genes obtained from limma–voom on the moderated t-statistics. Hallmark and MLL-fusion target gene sets were obtained from MSigDB database and gene sets for chromosomal positions were generated based on genes expressed in hotspots of transcriptomic changes (HTCs, see next paragraph). Heatmap of AML-related genes was generated by using aheatmap function in R on kallisto transcript per million values of genes among the top 300 in AML-related GeneCard list (see supplemental data) or having FC > 8.

Integrative chromosome plots

Start sites of each genes were retrieved from BioMart (https://useast.ensembl.org/biomart) and were used to plot the log2 (FC) of DEGs, obtained from limma–voom analyses, along each chromosome. The normalized average log2(FC) was computed based on log2(FC) from the entire list of expressed genes generated by limma–voom analysis. These genes were sorted

according to their genomic position and each chromosome was sliced into 25 equal regions. For each region, the average log2 (FC) has been computed. Each average has then been multiplied by the number of upregulated genes (if average > 0) or downregulated genes (if average < 0) and finally divided by the total number of genes of the region. Normalized average log2(FC) were plotted at the center position of each of the 25 regions. Based on the empirical observation that the normalized average log2 (FC) was higher than 0.45 at the LOH position in p arm of chromosome 6, HTC regions were established when more than two (sub 25) contiguous regions had normalized average log2 (FC) above 0.45 or below −0.45. The positions of LOH and allelic ratio aberrations were identified by eSNP-karyotyping analysis. Plots were built with Graphpad v7.00.

eSNP-karyotyping

Detection of chromosomal aberrations using RNA-seq data was performed using eSNP karyotyping as previously described.[16]

MHC-I expression analysis

Quantification of MHC class I, HLA-A, B, C surface expression was performed using the QIFIKIT bead-based flow cytometric assay (Dako, Agilent Technologies, Santa-Clara, CA, K0078) according to the manufacturer's recommendations. Briefly, triplicates of THP-1 ATCC and DSMZ samples were incubated first with an Fc-receptor blocking reagent (anti-CD32; Becton Dickinson (BD), Bedford, MA, 552930) for 10 min at room temperature, followed by incubation with either HLA-ABC-FITC (Invitrogen, Carlsbad, CA, 14-9983-82) or Mouse IgG2a κ Isotype Control (BD 553454) for 30 min at 4°C. Secondary staining was performed using FITC-conjugated antibody provided in the kit. Analysis was made by flow cytometry (BD FACS Canto II) and number of HLA molecules was determined by interpolation on the standard curve of the bead populations.

For quantification of HLA-A surface expression, triplicates of THP-1 ATCC and DSMZ samples were incubated with anti-CD32 and then either with anti-HLA-A2-PE (BD 558570) or isotype IgG2b,k-PE (BD 559529) for 20 min at 4°C. MFI of samples were analyzed using flow cytometry and data were analyzed with FlowJo software v10.5.3 (Tree Star Inc., Ashland, OR). Analyses were made with Graphpad v7.00.

Morphological assessment of ATCC and DSMZ cells

Cells were seeded at 0.2 million cells ml^{-1} in six-well plates and imaged 48 hr after seeding using Leica DMIRB microscope and QCapture software.

Data availability

Raw RNA-Seq data are openly available through Gene Expression Omnibus (GEO GSE130985). Other data can be made available upon reasonable request.

Results and Discussion

THP-1 cells, freshly purchased from ATCC or DSMZ, were submitted for STR profiling to the ATCC cell line authentication service. STR profiling, considered gold standard for cell line authentication testing, is based on examining the varying number of polymorphic repeats and comparing them with size standards.[17] The percentage match between two cell lines can be influenced by the algorithm used.[8] In the present case, the percentage match was calculated based on 8 core STR loci plus amelogenin (Supporting Information Table S1). ATCC authentication service used the Masters algorithm (= number of shared alleles between query sample and ATCC database profile/total number of alleles in the ATCC database profile) which yielded a 100% match [(15/15) × 100 = 100%]. However, the Tanabe algorithm [= (2 × number of shared alleles)/(number of alleles in sample#1 and number of alleles in sample#2)] yielded a 93.75% match: [(2 × 15)/(15 + 17)] × 100 = 93.75%. The discrepancy highlighted by the Tanabe algorithm is due to the fact that the THP-1 cell line from ATCC shows loss of two STR alleles (D13S317 and vWA loci). In accordance with this, the ATCC and DSMZ reference STR profiles of the THP-1 cell line found in the Cellosaurus database[18] show slight differences. Hence, THP-1 cells from ATCC and DSMZ are two different cell lines, one of them presenting at least a minor genetic drift. Additionally, based on STR profiling data from Cellosaurus database, we found that of 12 THP-1 sources reported, 6 had an exact STR match with the ATCC profile observed here while 2 had an exact match with the DSMZ profile (Supporting Information Table S2). In total, we identified at least four variants of the THP-1 cell line based on the profiling data (Supporting Information Table S2). In our case, since THP-1 cells are microsatellite stable, we observed that the modified STR profile was exclusively due to loss of STR alleles; in contrast to MSI+ cell lines in which alterations in STR profiles are due to both occurrence and loss of new and existing alleles, respectively.[19] We next sought to investigate the extent of the divergence between the two ATCC and DSMZ cell lines as well as its possible biological consequences. To address this, we first examined the HLA type of each cell line using both next-generation DNA sequencing and PCR-SSP. While the HLA type of ATCC cells was identical to the one found in a previous report (HLA-A*02:01/02:01 and HLA-B*15:11/15:11),[20] DSMZ cells were heterozygous for both HLA-A and HLA-B alleles (HLA-A*02:01/24:02 and HLA-B*15:11/35:01; Supporting Information Table S3 and Fig. S1). To confirm this observation, we performed RNA-sequencing of both cell lines and determined their HLA type with the Optitype software.[21] We also performed this analysis on publicly available RNA-Seq data on DSMZ THP-1 cells.[22] Concordant results were obtained with all methods: at both *HLA-A* and *HLA-B* loci, THP-1 cells from ATCC had a single allele whereas DSMZ THP-1 cells had two. We concluded that ATCC cells have undergone LOH before integration or during maintenance in this repository. Interestingly, in the original

study describing the establishment of THP-1, the HLA type was based on serological typing and reported as HLA-A2, A9, B5 (nomenclature of now-obsolete HLA typing system).[3] The apparent discrepancy between this HLA serotyping and the HLA genotype of DSMZ THP-1 can be explained by the lower resolution of serological typing and HLA antisera cross-reactivity.[23] Indeed, HLA-B5 antisera has been shown to be cross-reactive to HLA-B35 and HLA-B15 (HLA-B alleles identified in DSMZ THP-1).[20,24] Moreover, based on DNA sequencing and HLA-A9 antisera reactivity, A*24:02 is now classified under HLA-A9 antigen family (along with HLA A*23:01 and A*24:03).[25] Finally, A*24:02 being the most common *HLA-A* allele in Japan, its presence in DSMZ cells is consistent with the reported ethnic origin of THP-1 cells.[3] Altogether, these observations suggest that DSMZ THP-1 cells are more similar to the original cells described in 1980 than ATCC's.

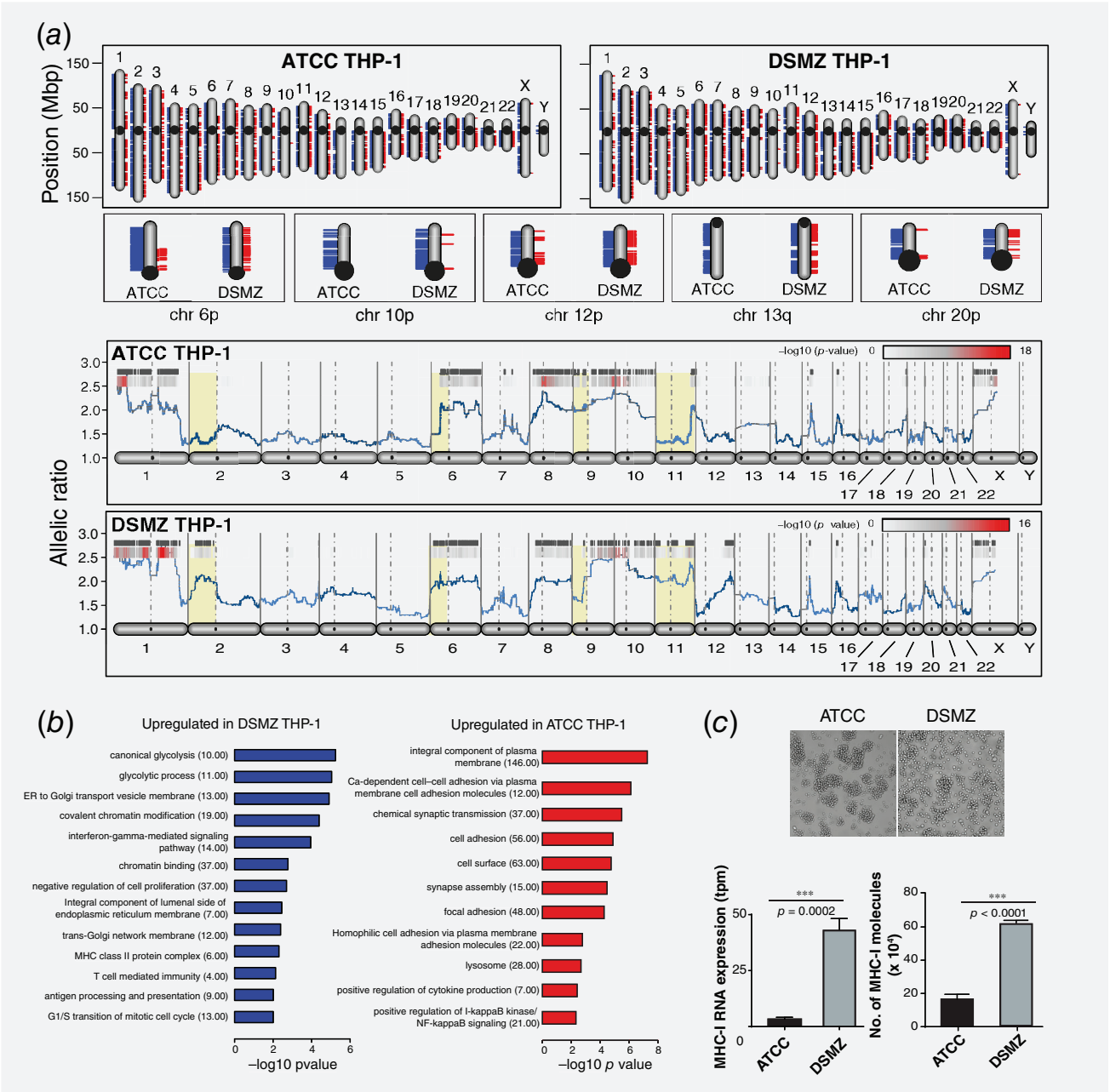

Figure 1. ATCC and DSMZ THP-1 cells show different cytogenetic aberrations and transcriptomic profiles. (*a*) *Top panel*: LOH analysis by eSNP-karyotyping. Blue—homozygous and red—heterozygous SNPs. *Bottom panel*: Allelic ratio analysis (differences between ATCC and DSMZ cells are highlighted in light yellow). (*b*) DAVID annotation of upregulated DEGs in DSMZ and ATCC cells. (*c*) *Top panel*: Light microscopy images after 48 hr of seeding at identical cell densities. *Bottom panel*: MHC-I expression at RNA level from limma–voom analysis (left) and protein level by flow cytometry (right: total HLA-ABC molecule number assessed by bead-based assay). Statistical significance was determined using unpaired *t*-test.

Using eSNP-karyotyping,[16] we confirmed LOH of the p arm of chromosome 6 (location of the HLA complex) of ATCC cells (Fig. 1*a*, top panel), consistent with a reported karyotypic analysis of ATCC THP-1.[26] Unexpectedly, our eSNP-karyotyping further provided evidence of LOH specific to ATCC cells in chromosomes 10p, 12p, 13q and 20p. Analysis of allelic ratios (major/minor SNPs), which highlights uneven chromosome copy numbers (Supporting Information Fig. S2), showed similar patterns in cell lines for most chromosomes, with major differences present in chromosomes 2p, 6p, 9p and 11 (Fig. 1*a*, bottom panel). We next performed a differential gene expression analysis to evaluate the potential

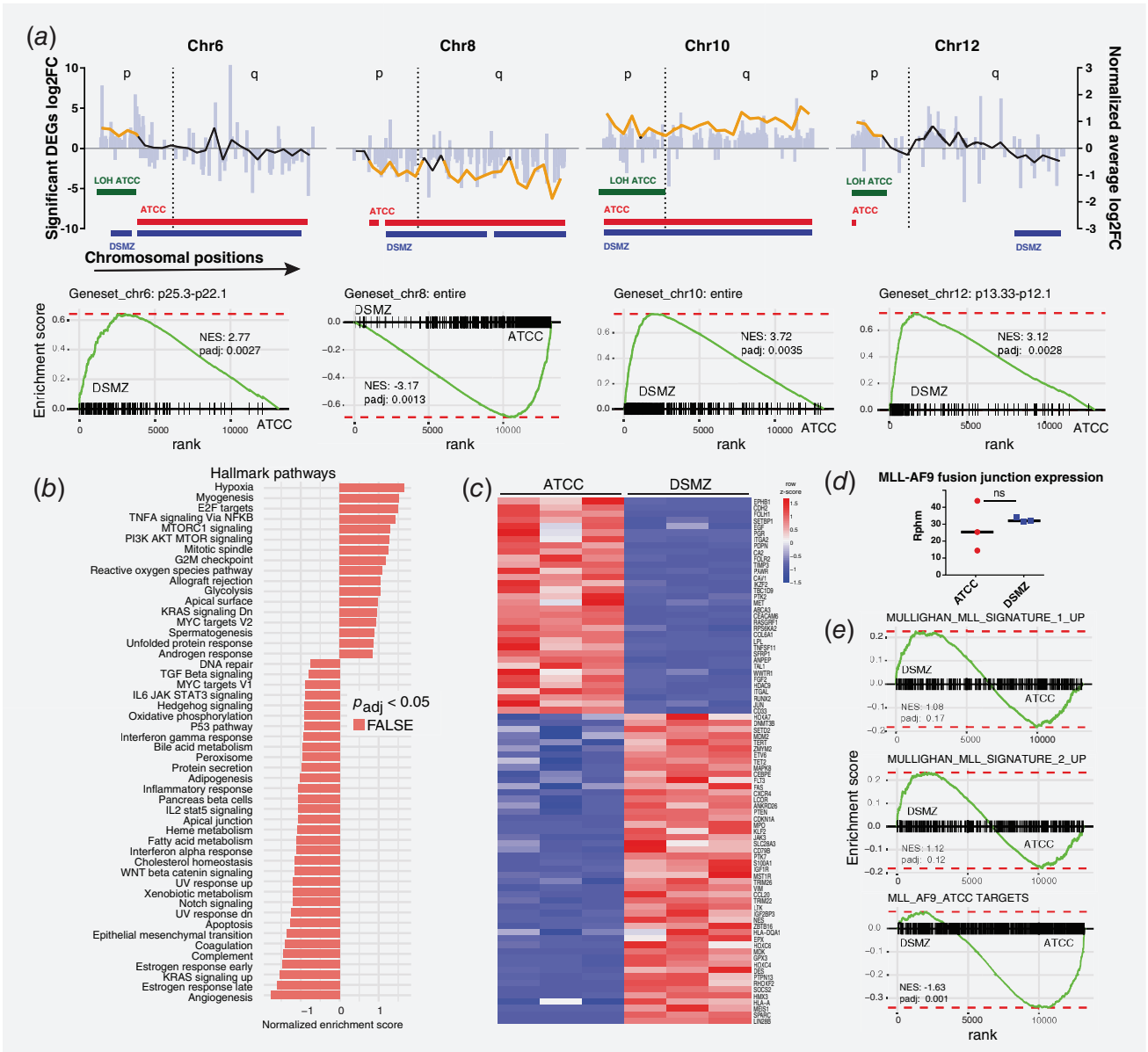

Figure 2. Divergence in cytogenetic aberrations leads to major transcriptomic differences between ATCC and DSMZ cells. (*a*) *Top panel*: Integrative chromosome plots depicting log2(FC) of significant DEGs (light blue bars) at their gene start position (*y*-axis) and normalized average log2(FC) in RNA expression (black lines). HTCs are indicated in orange. Green bars represent LOH regions in indicated cells, while horizontal red and blue bars represent aneuploidy regions (eSNP-karyotyping data). *Bottom panel*: GSEA of genes located in the indicated HTCs. (*b*) Normalized enrichment scores for hallmark gene sets evaluating enrichment of biological processes. (*c*) Heatmap of AML-related genes among the DEGs. Genes are sorted by their log2FC. (*d*) MLL-AF9 fusion junction expression in ATCC and DSMZ cells (count of fusion reads, normalized to the total number of reads); Rphm, reads per hundred million. Statistical significance was determined using unpaired *t*-test; ns, nonsignificant. (*e*) GSEA of gene sets associated with MLL fusion AML, established from studies using patient specimens (top two panels) or THP-1 cells (bottom panel); NES, normalized enrichment score; padj, false discovery rate adjusted *p*-value.

impact of these genomic discrepancies. This revealed that the two cell lines had markedly different transcriptomes as evidenced by principal component analysis (PCA) and volcano plot (2,627 differentially expressed genes [DEGs]; Supporting Information Fig. S3*a* and Table S4). These alterations likely have pervasive impact on numerous biological processes, as evidenced by gene ontology (GO) term analysis (Fig. 1*b* and Supporting Information Tables S5 and S6). They also modify cell phenotype since DSMZ cells expressed threefold to fourfold more MHC-I molecules (either total HLA or HLA-A*02, an allele shared by both cell lines) whereas ATCC cells formed more numerous aggregates, in agreement with their overexpression of genes linked to cell adhesion GO terms (Fig. 1*c* and Supporting Information Fig. S3*b*).

To establish the causative relationship between the cytogenetic and transcriptomic alterations, we combined our eSNP and DEG data for each chromosome (Fig. 2*a*, top panel, Supporting Information Fig. S4 and Tables S7 and S8). This revealed that in specific chromosomes (8, 9, 10, 12 and X), more than 20% of expressed genes were DEGs (respectively 40.2, 24.7, 50.1, 23.1 and 37.3%). To evaluate the connection between genomic location and gene expression, we computed a normalized average log2(Fold Change [FC]; Materials and Methods) along each chromosome. This allowed us to identify several HTCs between both cell lines. Strikingly, the positions of these HTCs correlated well with those of LOH or allelic ratio aberrations and their normalized average log2(FC) was either in agreement with eSNP analysis (e.g., the specific LOH in chr6p and chr12p of ATCC cells resulted in overexpression of these regions by DSMZ cells) or provided further insights into it (e.g., copy number of chromosomes 8 and 10 could be higher in ATCC and DSMZ cells, respectively). Finally, we collected the list of genes located in the HTCs (chr1:q24.3–q42.13, chr6:p25.3–p22.1, chr8, chr9:q12–q21.31, chr10, chr12:p13.33–p12.1, chr20:p13–p11.21 and chrX) and compared their expression using the GSEA tool (Fig. 2*a*, bottom panel, Supporting Information Figure S5 and Table S9). As controls, we examined various biological processes with the HALLMARK genes sets, which were constructed independently of gene chromosomal location[27] in contrast to our HTCs (Fig. 2*b*). This showed that every HTC gene set was significantly differentially expressed, in contrast with controls. Altogether, these results show that transcriptomic and phenotypic differences are closely linked to differential chromosomal aberrations between both cell lines.

Finally, we tested whether inter-cell line divergences could impact the study of genes typically associated with AML. We extracted the full list of AML-related genes from the GeneCards database and queried each one of them in our list of DEGs. This showed that 378/2646 AML-related genes were among the DEGs, 84 being either differentially expressed more than eightfold or among the top 300 genes of the list (Fig. 2*c* and Supporting Information Table S10). Among the four DEGs with the highest log2(FC) (>7) were *LIN28B*, *MEIS1* and *SPARC*, whose leukemic function has been documented in THP-1 cells.[28–30] Importantly, *MEIS1* is systematically overexpressed in AMLs harboring *MLL* fusions, where its expression is promoted by the fusion proteins and is implicated in leukemogenesis.[31] We therefore examined the expression of *MLL* fusion junction in both cell lines and found no difference in expression (Fig. 2*d*). GSEA of *MLL* fusion targets[32] were also not significant (Fig. 2*e*, top two panels). However, a gene set of *MLL* fusion targets established based on THP-1 cells[28] was significantly overexpressed in ATCC cells, showing that results obtained by studies of *MLL* fusion in THP-1 could differ based on their biorepository of origin (Fig. 2*e*, bottom panel and Supporting Information Table S11).

In summary, we demonstrate that THP-1 cells obtained from two different well-established biorepositories are in fact THP-1 variants that have undergone genetic drift and present important molecular and phenotypic differences. We demonstrate that these alterations can have deleterious effects on the reproducibility and conclusions of studies using this cell line as a model. The genetic heterogeneity and the unstable nature of tumor samples and cell lines, and their influence on tumor evolution and phenotypic variability, have been well reported in the genetically unstable HeLa and MCF-7 cells.[11,12] Specifically, MCF-7 cells from the same repository cultured simultaneously by two different labs quickly showed genetic differences and phenotypic variability.[33] In the present study, although THP-1 cells are considered as genetically stable, we found that freshly acquired cells from two different well-established cell banks (which are commonly assumed to provide rigorously identical cell lines across the world) are not the same entity due to genetic drift. This emphasizes the crucial necessity of clearly naming the source of each cell line used in research articles. Finally, our data also demonstrate that seemingly minor discrepancies in standard STR profiles (8 loci and amelogenin) should alert to the need for more in-depth evaluation because they may be the sign of extensive genetic drift with dramatic functional consequences. The primary purpose of STR testing in research was to evaluate cell line cross-contaminations. For this purpose, testing 8–16 STR loci is sufficient. However, this is much less effective for detection of the genetic drift in a given cell line. Increasing the number of STR loci tested (from 8–16 to 80–160 loci[34,35]) can substantially improve the accuracy of genetic drift (and LOH in particular) detection.[8] Evaluation of the functional importance of genetic drift may then be achieved by functional assays and RNA-sequencing.

Acknowledgements

We thank the following members of IRIC core facilities for technical assistance: S. Boissel from the genomic platform and P. Gendron from the bioinformatics platform. This work was supported by The Leukemia & Lymphoma Society of Canada.

Conflict of interest

The authors declare that they have no conflict of interest.

References

1. Krivtsov AV, Armstrong SA. MLL translocations, histone modifications and leukaemia stem-cell development. *Nat Rev Cancer* 2007;7:823–33.
2. Barabe F, Kennedy JA, Hope KJ, et al. Modeling the initiation and progression of human acute leukemia in mice. *Science* 2007;316:600–4.
3. Tsuchiya S, Yamabe M, Yamaguchi Y, et al. Establishment and characterization of a human acute monocytic leukemia cell line (THP-1). *Int J Cancer* 1980;26:171–6.
4. Baker M. 1,500 scientists lift the lid on reproducibility. *Nature* 2016;533:452–4.
5. Fusenig NE, Capes-Davis A, Bianchini F, et al. The need for a worldwide consensus for cell line authentication: experience implementing a mandatory requirement at the international journal of cancer. *PLoS Biol* 2017;15:e2001438.
6. MacLeod RA, Dirks WG, Matsuo Y, et al. Widespread intraspecies cross-contamination of human tumor cell lines arising at source. *Int J Cancer* 1999;83:555–63.
7. Capes-Davis A, Theodosopoulos G, Atkin I, et al. Check your cultures! A list of cross-contaminated or misidentified cell lines. *Int J Cancer* 2010;127:1–8.
8. Capes-Davis A, Reid YA, Kline MC, et al. Match criteria for human cell line authentication: where do we draw the line? *Int J Cancer* 2013;132:2510–9.
9. Drexler HG, Dirks WG, MacLeod RA, et al. False and mycoplasma-contaminated leukemia-lymphoma cell lines: time for a reappraisal. *Int J Cancer* 2017;140:1209–14.
10. Korch C, Spillman MA, Jackson TA, et al. DNA profiling analysis of endometrial and ovarian cell lines reveals misidentification, redundancy and contamination. *Gynecol Oncol* 2012; 127:241–8.
11. Liu Y, Mi Y, Mueller T, et al. Multi-omic measurements of heterogeneity in HeLa cells across laboratories. *Nat Biotechnol* 2019;37:314–22.
12. Ben-David U, Siranosian B, Ha G, et al. Genetic and transcriptional evolution alters cancer cell line drug response. *Nature* 2018;560:325–30.
13. Bosshart H, Heinzelmann M. THP-1 cells as a model for human monocytes. *Ann Transl Med* 2016;4:438.
14. Olarerin-George AO, Hogenesch JB. Assessing the prevalence of mycoplasma contamination in cell culture via a survey of NCBI's RNA-seq archive. *Nucleic Acids Res* 2015;43:2535–42.
15. Sergushichev AA. An algorithm for fast preranked gene set enrichment analysis using cumulative statistic calculation. *bioRxiv* 2016;2016:060012.
16. Weissbein U, Schachter M, Egli D, et al. Analysis of chromosomal aberrations and recombination by allelic bias in RNA-Seq. *Nat Commun* 2016;7:12144.
17. Butler JM. Genetics and genomics of core short tandem repeat loci used in human identity testing. *J Forensic Sci* 2006;51:253–65.
18. Bairoch A. The Cellosaurus, a cell-line knowledge resource. *J Biomol Tech* 2018;29:25–38.
19. Parson W, Kirchebner R, Muhlmann R, et al. Cancer cell line identification by short tandem repeat profiling: power and limitations. *FASEB J* 2005;19:434–6.
20. Battle R, Poole K, Haywood-Small S, et al. Molecular characterisation of the monocytic cell line THP-1 demonstrates a discrepancy with the documented HLA type. *Int J Cancer* 2013;132:246–7.
21. Szolek A, Schubert B, Mohr C, et al. OptiType: precision HLA typing from next-generation sequencing data. *Bioinformatics* 2014;30:3310–6.
22. Marzotto M, Bonafini C, Olioso D, et al. *Arnica montana* stimulates extracellular matrix gene expression in a macrophage cell line differentiated to wound-healing phenotype. *PLoS One* 2016;11:e0166340.
23. Howell WM, Carter V, Clark B. The HLA system: immunobiology, HLA typing, antibody screening and crossmatching techniques. *J Clin Pathol* 2010; 63:387–90.
24. Hein J, Bottcher K, Grundmann R, et al. Low resolution DNA typing of the HLA-B5 cross-reactive group by nested PCR-SSP. *Tissue Antigens* 1995;45:27–35.
25. Fussell H, Thomas M, Street J, et al. HLA-A9 antibodies and epitopes. *Tissue Antigens* 1996;47:307–12.
26. Odero MD, Zeleznik-Le NJ, Chinwalla V, et al. Cytogenetic and molecular analysis of the acute monocytic leukemia cell line THP-1 with an MLL-AF9 translocation. *Genes Chromosomes Cancer* 2000;29:333–8.
27. Liberzon A, Birger C, Thorvaldsdottir H, et al. The molecular signatures database (MSigDB) hallmark gene set collection. *Cell Syst* 2015;1:417–25.
28. Prange KHM, Mandoli A, Kuznetsova T, et al. MLL-AF9 and MLL-AF4 oncofusion proteins bind a distinct enhancer repertoire and target the RUNX1 program in 11q23 acute myeloid leukemia. *Oncogene* 2017;36:3346–56.
29. Alachkar H, Santhanam R, Maharry K, et al. SPARC promotes leukemic cell growth and predicts acute myeloid leukemia outcome. *J Clin Invest* 2014;124:1512–24.
30. Zhou J, Bi C, Ching YQ, et al. Inhibition of LIN28B impairs leukemia cell growth and metabolism in acute myeloid leukemia. *J Hematol Oncol* 2017;10:138.
31. Kumar AR, Li Q, Hudson WA, et al. A role for MEIS1 in MLL-fusion gene leukemia. *Blood* 2009; 113:1756–8.
32. Mullighan CG, Kennedy A, Zhou X, et al. Pediatric acute myeloid leukemia with NPM1 mutations is characterized by a gene expression profile with dysregulated HOX gene expression distinct from MLL-rearranged leukemias. *Leukemia* 2007;21:2000–9.
33. Kleensang A, Vantangoli MM, Odwin-DaCosta S, et al. Genetic variability in a frozen batch of MCF-7 cells invisible in routine authentication affecting cell function. *Sci Rep* 2016;6:28994.
34. Mizutani M, Yamamoto T, Torii K, et al. Analysis of 168 short tandem repeat loci in the Japanese population, using a screening set for human genetic mapping. *J Hum Genet* 2001;46:448–55.
35. Weaver DA, Hei TK, Hukku B, et al. Localization of tumor suppressor gene candidates by cytogenetic and short tandem repeat analyses in tumorigenic human bronchial epithelial cells. *Carcinogenesis* 2000;21:205–11.