Bioinformatics, 33(23), 2017, 3691–3695
doi: 10.1093/bioinformatics/btx502
Advance Access Publication Date: 7 August 2017
Discovery Note

OXFORD

Gene expression

Metagene projection characterizes GEN2.2 and CAL-1 as relevant human plasmacytoid dendritic cell models

Pedro Carmona-Sáez[1,*], Nieves Varela[2], María José Luque[2], Daniel Toro-Domínguez[1,2], Jordi Martorell-Marugan[1], Marta E. Alarcón-Riquelme[2,3] and Concepción Marañón[2,*]

[1]Bioinformatics Unit and [2]Genomic Medicine Department, GENYO, Centre for Genomics and Oncological Research, Pfizer/University of Granada/Andalusian Regional Government, PTS 18016, Granada, Spain and [3]Institute for Environmental Medicine, Karolinska Institutet, Stockholm, Sweden

*To whom correspondence should be addressed.

Associate Editor: Bonnie Berger

Received on August 13, 2016; revised on July 4, 2017; editorial decision on August 3, 2017; accepted on August 6, 2017

Abstract

Motivation: Plasmacytoid dendritic cells (pDC) play a major role in the regulation of adaptive and innate immunity. Human pDC are difficult to isolate from peripheral blood and do not survive in culture making the study of their biology challenging. Recently, two leukemic counterparts of pDC, CAL-1 and GEN2.2, have been proposed as representative models of human pDC. Nevertheless, their relationship with pDC has been established only by means of particular functional and phenotypic similarities. With the aim of characterizing GEN2.2 and CAL-1 in the context of the main circulating immune cell populations we have performed microarray gene expression profiling of GEN2.2 and carried out an integrated analysis using publicly available gene expression datasets of CAL-1 and the main circulating primary leukocyte lineages.

Results: Our results show that GEN2.2 and CAL-1 share common gene expression programs with primary pDC, clustering apart from the rest of circulating hematopoietic lineages. We have also identified common differentially expressed genes that can be relevant in pDC biology. In addition, we have revealed the common and differential pathways activated in primary pDC and cell lines upon CpG stimulatio.

Availability and implementation: R code and data are available in the supplementary material.

Contact: pedro.carmona@genyo.es or concepcion.maranon@genyo.es

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Plasmacytoid dendritic cells (pDC) are key players in the control of the immune responses. They are located at the interface between the innate and adaptive branches of the immune system, where they control both immunity and tolerance (Mathan *et al.*, 2013). pDC are main actors in the stimulation of B lymphocytes (Shaw *et al.*, 2010), NK cells (Swiecki *et al.*, 2010) and conventional dendritic cells (cDCs). Therefore the study of this cell line is crucial for the understanding of protective responses against pathogens or tumors and mechanisms of deregulated responses in autoimmunity.

pDC derive from a common dendritic cell precursor (Satpathy *et al.*, 2012) in the bone marrow, in steady state, are located mainly in the marginal zone and the T-cell areas of the human spleen (Nascimbeni *et al.*, 2009). They are found at extremely low frequencies in the blood, and they do not proliferate and die in less than 72 h in culture (Sisirak *et al.*, 2011). In contrast to cDCs, there is no

standardized protocol to generate *bona fide* pDC from hematopoietic precursors *in vitro* and hence the study of human pDC biology is a great challenge to immunologists.

In the last few years, a renewed interest in the leukemic counterparts of pDC has arisen and some pDC-like lines have been generated from biopsies of pDC neoplasm patients (Chaperot *et al.*, 2001; Watanabe *et al.*, 2010). Several authors have reported their similarities with primary pDC in terms of surface phenotype (Chaperot *et al.*, 2006) and cytokine responses (Wang *et al.*, 2016) or gene regulation (Cheng *et al.*, 2015) and therefore they are promising models for the study of the pDC biology. However, there is lack of more systematic studies to support the use of these cell lines as relevant pDC models. Therefore, a more global characterization of these cell lines will benefit our understanding of their biology and will help us evaluate how well they represent *bona fide* pDC.

The main goal of this work was to characterize the global gene expression profile of the most commonly used pDC models, the GEN2.2 and CAL-1 leukemic cell lines, and to compare them with gene expression signatures derived from a panel of immune cell populations. To accomplish this, we generated microarray gene expression data of GEN2.2 cells and performed an integrated analysis with previously reported datasets of CAL-1 and different primary hematopoietic cell types. Using public gene expression data we defined the main immune system-related metagenes covering different cell populations, based on the methodology proposed by Tamayo *et al.* (2007). Projection of CAL-1 and GEN2.2 expression profiles into this metagene space allowed us to define the main similarities of these cell lines to other immune cell populations and to establish a clear connection with pDC.

Additionally, we also analyzed microarray data from GEN2.2, CAL-1 and pDC after stimulation with a synthetic oligonucleotide containing unmethylated CpG motifs. This analysis revealed significant similarities in the molecular pathways that were induced in GEN2.2, CAL-1 and pDCs after stimulation, which provides additional evidence to support the use of these cell lines as relevant pDC models.

2 Materials and methods

2.1 Microarray gene expression profiling

GEN2.2 cell line was grown at log phase in complete medium + 10% calf fetal serum. CpG stimulation was carried out adding 1 μM of ODN2006 (Life Technologies) for 2 h in triplicates. Total RNA was isolated from treated and untreated samples with High Pure RNA Isolation Kit (Roche Diagnostics, Mannheim, Germany) and 300 ng was processed using Illumina TotalPrep RNA Amplification Kit and hybridized with Illumina Human-HT12 v4 expression beadchip. Raw data were exported from Illumina GenomeStudio and processed using the normexp followed by quantile normalization. Gene expression values for replicated genes were aggregated using the median value. Gene expression data is available in NCBI-GEO with accession number GSE59887.

2.2 Analysis of public data and defining immune system metagenes

NCBI-GEO database (Barrett *et al.*, 2013) was queried to find pDC-related datasets. After careful revision six gene expression datasets were selected for the study (Supplementary Table S1 in additional file 1). Cel files from affymetrix platforms were downloaded and processed applying RMA and quantile normalization. Samples were inspected for quality control and sample mDC 16 from the dataset GSE28490 was removed from the analysis as this was not correlated with the mDC specific metagene (data not shown). Soft-formatted files from GEO were used for Illumina and NCI/ATC Hs-Operon V3. The median was computed for probe sets corresponding to the same gene, non annotated probes were discharged and gene expression data were rank normalized.

The dataset generated by Allantaz *et al.* (2012), which contains gene expression profiles of different human immune cell subsets, was used to define immune system related metagenes using the methodology reported by Tamayo *et al.* (2007). Briefly, Non-Negative Matrix Factorization (NMF) was applied to reduce the dimensionality and define a small subset of factors that summarizes the main features of the cell populations. These features represent subsets of genes co-expressed in subsets of experimental conditions (Carmona-Saez *et al.*, 2006). The number of classes (nine cell populations) was known, but to evaluate the best factorization rank we computed different stability metric, including the correlation cophenetic coefficient (ccc) for different values of k (Brunet *et al.*, 2004). Based on this analysis we selected $k = 9$ as the best factorization rank (Supplementary Fig. S1 in additional file 1), which is in agreement with the number of different cell populations.

Samples from independent datasets were projected into the metagene space by means of pseudoinverse of factor matrix after rank normalization and matching datasets by common genes. Hierarchical clustering analysis was then applied to group samples based on similarities of metagene profiles. All analyses were performed in the R statistical environment. All processed data and R scripts are available in the additional material (additional File 5).

2.3 Differential expression and functional analysis

To define a common signature among pDCs and immune cell lines we compared each pDCs related-cell (primary pDCs, GENE2.2 and CAL-1) with all immune cell population. Linear models were applied for differential expression analysis and genes that were found significant in all comparisons (fdr corrected P-value < 0.01 and absolute log fold change > 0.5) were selected. For differential expression analysis in stimulated cells we used RankProducts which is particularly powerful for a small number of replicates (Breitling *et al.*, 2004). Enrichment analysis in gene lists was performed with the GeneCodis (Carmona-Saez *et al.*, 2007) and Gene Set Enrichment Analysis (GSEA).

3 Results

3.1 Metagene projection reveals that GEN2.2 and CAL-1 cell lines share common gene expression patterns with pDC

In order to create a projection space with immune system-related metagenes we used as a model the dataset generated by Allantaz *et al.* (2012). This study provides gene expression data for the main immune cell populations and, additionally, an independent dataset from a set of different samples. Therefore, this study provides a gold standard for our analysis, enabling us to extract metagenes that summarize the gene expression programs of different immune cell populations. In addition, the second set of samples that can be used as internal controls of the data analysis pipeline.

Once the metagenes were defined, samples from different datasets were projected into this new space, including GEN2.2, CAL-1 (Cisse *et al.*, 2008), the validation dataset mentioned before, and other public datasets that contained gene expression profiles of primary pDC (Supplementary Table S1 in additional file 1).

Specifically, we used data from a study carried out by Haniffa *et al.* (2012) in which gene expression profiles of pDC, monocytes and dendritic cell subpopulations were compared. Initially, we expected that each metagene would define the distinguishing features of each subclass, but a careful exploration of the results revealed that metagene 6 was associated with two different cell populations, $CD4^+$ and $CD8^+$ T-cells, evidencing that both cell lines share some common gene expression features. This is in agreement with the fact that $CD4^+$ and $CD8^+$ T lymphocytes constitute the most related lineages in this dataset. Indeed, these two cell subpopulations were considered as a single group of 'T cells' in the original work because few transcripts were exclusive of $CD4^+$ or $CD8^+$ cells (Allantaz *et al.*, 2012). On the other hand, metagene 1 presented features that were partially associated with different samples in the dataset, rather than capturing specific features of one particular class. The rest of metagenes revealed specific features for the different cell populations. Metagene 2 and 3 were specific to eosinophils and neutrophils respectively. Metagenes 4, 5, 7, 8 and 9 showed clear specificity for monocytes, NK cells, B cells, cDC (named originally mDC) and pDC respectively (Fig. 1A). As can be noted some samples from $CD8^+$ T cells also showed high coefficients for factor 5 (NK cells), likely as a reflection of their shared functional features. A GSEA analysis was performed in each metagene and we found that genesets derived from the comparison of pDC and other cell lines were highly associated to metagene 9 (Supplementary Table S2 in additional file 1), which reinforces the notion that this metagene captured the main patterns associated to pDC.

As expected, samples from the validation dataset (Fig. 1C) showed high coefficients for the corresponding metagenes extracted from the model dataset. The projection of the other primary cell dataset (Fig. 1E) fits also perfectly with the expected model. The three monocyte subsets ($CD14^+$, $CD16^+$ and $CD14^+CD16^+$) shared high coefficients for metagene 4, while the two cDC populations ($CD141^+$ and $CD1c^+$) showed high values for metagene 8, and pDC gave high coefficients for metagene 9. Moreover, the pDC neoplasm cell lines CAL-1 (Fig. 1D) and GEN2.2 (Fig. 1B) showed the highest coefficients in the pDC metagene. These data suggest that the GEN2.2 and CAL-1 cell lines share common gene expression programs with primary pDC.

To further explore the similarity among cell types based on gene expression programs a hierarchical clustering analysis was performed. Using all genes the different samples were grouped based on the platform/study rather than distinguishing different cell lineages (Fig. 2A). On the other hand, clustering analysis in the metagene space grouped identical or nearly related primary cell types (Fig. 2B). All NK, B lymphocytes, pDC, cDC, monocytes, neutrophils and eosinophils clustered in discrete branches regardless of the platform or dataset. cDC (both $CD141^+$ and $CD1c^+$) and pDC clustered in related branches in agreement with their common origin (Satpathy, 2012). Regarding the cell lines, both GEN2.2 and CAL-1 cells clustered with all the pDC samples included in the study. See also Supplementary Figure S2 in additional file 1. We assessed clustering stability by multiscale bootstrap resampling (Supplementary Figs S3 and S4 in additional file 1) and a consensus clustering approach (Supplementary Fig. S5 in additional file 1), finding an overall high stability.

3.2 Definition of the common signature among pDC, GEN2.2 and CAL-1

A total of 211 genes were found significant across all comparisons of pDC related cells (primary pDC, GEN2.2 and CAL-1) and the rest of immune cell populations (Supplementary Fig. S6 in additional file 1). Some genes previously related to the biology of pDC were included in this list, i.e. *SLC7A5*, *SRPX*, *COL24A1*, *CDH1*,

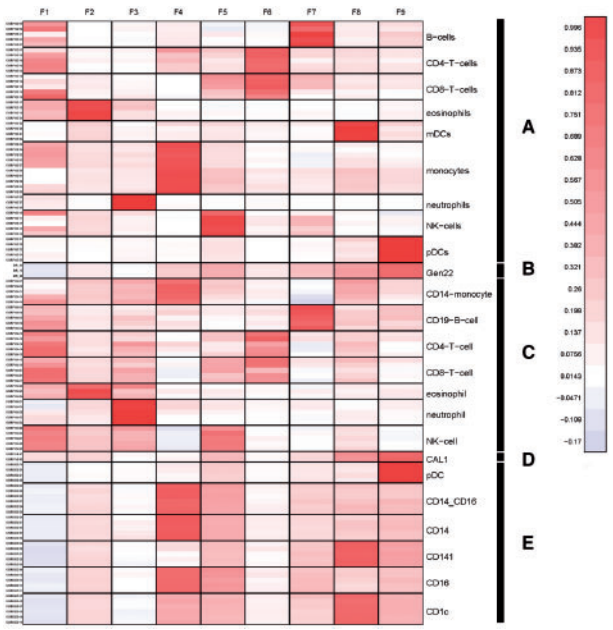

Fig. 1. Analysis of samples in the metagene space. Heatmap representing encoding coefficients of each sample (rows) in each metagene (columns) for model dataset (**A**) and test datasets (**B–E**). A: GSE28490; B: GSE59887 (GEN2.2); C: GSE28491; D: GSE12507 and E: GSE15215

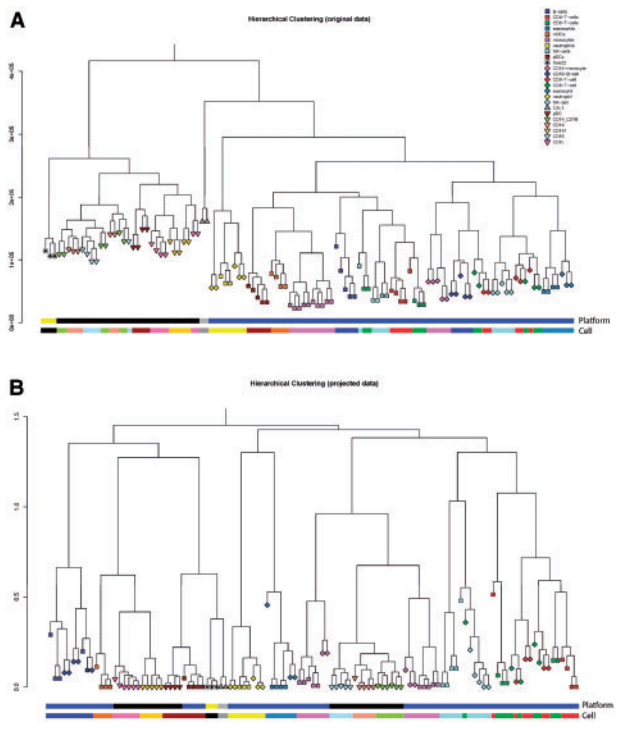

Fig. 2. Hierarchical clustering analysis. Hierarchical clustering analysis of samples of all merged and normalized datasets before (**A**) and after (**B**) metagene projection. Same shape represents samples from the same study and same color represents an equivalent cell population. Information about study and cells is also represented with the same color in bars below the dendrogram. GEN2.2 samples are marked with asterisks

CLEC4C, TNFRSF21, SCN9A, GPM6B, PACSIN1, NRP1, KCNK17, DAB2, SLC7A11, BSPRY, MCC, SLC35F3, MOXD1 or *AHI1* (Allantaz *et al.*, 2012), some of them also reported in independent works such as *CLEC4C* (Dzionek *et al.*, 2000), or *GPM6B* and *PACSIN1* (Robbins *et al.*, 2008). Many others, such as *SLITRK5*, or *SEC61A2* had not been described before as pDC-specific, although we found evidence of their pDC-expression pattern in other gene expression datasets (compiled in Immuno-Navigator: http://sysimm.ifrec.osaka-u.ac.jp/immuno-navigator/). The analysis of GO annotations revealed that functions related with transmembrane transport and cell adhesion were enriched in the list of over-expressed genes (see Supplementary Table S3 in additional file 1), processes that have not been specifically studied in the context of pDC biology. We also evaluated expression patterns of pDC related genes from previously published work (Supplementary Fig. S7 in additional file 1), and we found a good agreement. This comparison also revealed some discrepancies since some previously reported genes are over-expressed in other cell types. The comparison of the transcription profiles of pDC-like cell lines and primary pDC confirmed that the main pathways over-expressed are a consequence of the tumoral transformation, since they were all related with DNA replication and cell cycle (Supplementary Table S4 in additional file 1 and additional file 4). Also of note, pathways related with transcription regulation, signal transduction and intracellular traffic were consistently down-regulated in GEN2.2 and CAL-1 compared with pDC (see additional file 4).

3.3 CAL-1 and GEN2.2 activate common pathways with primary pDC after CpG stimulation

Oligonucleotides containing unmethylated CpG motifs stimulate pDC to produce type I interferons and proinflammatory cytokines (Kerkmann *et al.*, 2003). To determine whether CAL-1 and GEN2.2 mimic the response of primary pDC after stimulation, we compared the gene expression profiles of these cell lines after stimulation with CpG. To this end, public gene expression datasets of stimulated and non-stimulated samples of CAL-1 (Steinhagen *et al.*, 2012) and pDC (Loures *et al.*, 2015) were retrieved from GEO and processed as described in methods. Microarray gene expression data were experimentally generated for GEN2.2 after 2 h of CpG treatment. Differential expression analysis yielded 273 genes in GEN2.2, 111 in CAL-1 and 125 in primary pDC whose expression was significantly increased (corrected P-value <0.05) after stimulation (Supplementary Table S5 in additional file 1). Although the overlapping of genes in these three lists was moderate, a GSEA analysis revealed a significant enrichment of GEN2.2 and CAL-1 gene sets in the primary pDC signature (Supplementary Fig. S8 in additional file 1). In addition, functional analysis of GO terms also indicated that the majority of the biological processes that were significantly activated upon pDC stimulation were also found enriched in stimulated CAL-1 and GEN2.2 (Supplementary Fig. S9 in additional file 1). However, some pathways were under-represented in CAL-1 and GEN2.2, such as negative regulation of T cell differentiation and collagen catabolic processes. Other pathways were exclusively defective in stimulated CAL-1 cells.

4 Conclusion

Our data show that GEN2.2 and CAL-1 share common expression programs with primary human pDC. In addition, their responses to CpG genes also activate common pathways, supporting the use of both cell lines as pDC models *in vitro*. On the other hand, our study reveals that some pDC specific pathways are not fully recapitulated in GEN2.2 and CAL-1. All these similarities and differences should be taken into account when a particular pDC function needs to be investigated using these models.

Acknowledgements

We thank Laurence Chaperot for providing GEN2.2 cells, Luis Javier Martinez and Céline Coppard for their technical assistance and Pablo Tamayo for sharing the Metagene R library.

Funding

This work has been supported by the Instituto de Salud Carlos III (PI12/2558, PI10/0552, PI13/0522) in part through FEDER funds of the European Union, the Fundación Ramón Areces and by Spanish Ministry of Economy, Industry and Competitiveness (grant SAF2016-78631-P).

Conflict of Interest: none declared.

References

Allantaz,F. *et al.* (2012) Expression profiling of human immune cell subsets identifies miRNA-mRNA regulatory relationships correlated with cell type specific expression. *PloS One*, **7**, e29979.

Barrett,T. *et al.* (2013) NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res.*, **41**, D991–D995.

Breitling,R. *et al.* (2004) Rank products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments. *FEBS Lett.*, **573**, 83–92.

Brunet,J.-P. *et al.* (2004) Metagenes and molecular pattern discovery using matrix factorization. *Proc. Natl. Acad. Sci. USA*, **101**, 4164–4169.

Carmona-Saez,P. *et al.* (2006) Biclustering of gene expression data by non-smooth non-negative matrix factorization. *BMC Bioinformatics*, **7**, 78.

Carmona-Saez,P. *et al.* (2007) GENECODIS: a web-based tool for finding significant concurrent annotations in gene lists. *Genome Biol.*, **8**, R3.

Chaperot,L. *et al.* (2001) Identification of a leukemic counterpart of the plasmacytoid dendritic cells. *Blood*, **97**, 3210–3217.

Chaperot,L. *et al.* (2006) Virus or TLR agonists induce TRAIL-mediated cytotoxic activity of plasmacytoid dendritic cells. *J. Immunol. Baltim. Md. 1950*, **176**, 248–255.

Cheng,M. *et al.* (2015) Characterization of species-specific genes regulated by E2-2 in human plasmacytoid dendritic cells. *Sci. Rep.*, **5**, 10752.

Cisse,B. *et al.* (2008) Transcription factor E2-2 is an essential and specific regulator of plasmacytoid dendritic cell development. *Cell*, **135**, 37–48.

Davis,S. and Meltzer,P.S. (2007) GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics*, **23**, 1846–1847.

Dzionek,A. *et al.* (2000) BDCA-2, BDCA-3, and BDCA-4: three markers for distinct subsets of dendritic cells in human peripheral blood. *J. Immunol. Baltim. Md. 1950*, **165**, 6037–6046.

Haniffa,M. *et al.* (2012) Human tissues contain CD141hi cross-presenting dendritic cells with functional homology to mouse CD103+ nonlymphoid dendritic cells. *Immunity*, **37**, 60–73.

Kerkmann,M. *et al.* (2003) Activation with CpG-A and CpG-B oligonucleotides reveals two distinct regulatory pathways of type I IFN synthesis in human plasmacytoid dendritic cells. *J. Immunol. Baltim. Md. 1950*, **170**, 4465–4474.

Loures,F.V. *et al.* (2015) Recognition of *Aspergillus fumigatus* Hyphae by human plasmacytoid dendritic cells is mediated by dectin-2 and results in formation of extracellular traps. *PLoS Pathog.*, **11**, e1004643.

Maeda,T. *et al.* (2005) A novel plasmacytoid dendritic cell line, CAL-1, established from a patient with blastic natural killer cell lymphoma. *Int. J. Hematol.*, **81**, 148–154.

Mathan,T.S.M.M. *et al.* (2013) Human plasmacytoid dendritic cells: from molecules to intercellular communication network. *Front. Immunol.*, **4**, 372.

Nascimbeni,M. *et al.* (2009) Plasmacytoid dendritic cells accumulate in spleens from chronically HIV-infected patients but barely participate in interferon-alpha expression. *Blood*, **113**, 6112–6119.

Robbins,S.H. *et al*. (2008) Novel insights into the relationships between dendritic cell subsets in human and mouse revealed by genome-wide expression profiling. *Genome Biol*., **9**, R17.

Satpathy,A.T. *et al*. (2012) Re(de)fining the dendritic cell lineage. *Nat. Immunol*., **13**, 1145–1154.

Shaw,J. *et al*. (2010) Plasmacytoid dendritic cells regulate B-cell growth and differentiation via CD70. *Blood*, **115**, 3051–3057.

Sisirak,V. *et al*. (2011) CCR6/CCR10-mediated plasmacytoid dendritic cell recruitment to inflamed epithelia after instruction in lymphoid tissues. *Blood*, **118**, 5130–5140.

Steinhagen,F. *et al*. (2012) Activation of type I interferon-dependent genes characterizes the "core response" induced by CpG DNA. *J. Leukoc. Biol*., **92**, 775–785.

Swiecki,M. *et al*. (2010) Plasmacytoid dendritic cell ablation impacts early interferon responses and antiviral NK and CD8(+) T cell accrual. *Immunity*, **33**, 955–966.

Tamayo,P. *et al*. (2007) Metagene projection for cross-platform, cross-species characterization of global transcriptional states. *Proc. Natl. Acad. Sci. USA*, **104**, 5959–5964.

Wang,F. *et al*. (2016) Alarmin human α defensin HNP1 activates plasmacytoid dendritic cells by triggering NF-κB and IRF1 signaling pathways. *Cytokine*, **83**, 53–60.

Watanabe,N. *et al*. (2010) Transformation of dendritic cells from plasmacytoid to myeloid in a leukemic plasmacytoid dendritic cell line (PMDC05). *Leuk. Res*., **34**, 1517–1524.