Data in Brief 20 (2018) 1797–1803

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Dataset on the response of Hut78 cells to novel rexinoids

Jennifer Hackney Price*, Bentley J. Hanish, Carl E. Wagner, Ichiro Kaneko, Peter W. Jurutka, Pamela A. Marshall

School of Mathematical and Natural Sciences, Arizona State University, USA

ARTICLE INFO

Article history:
Received 26 April 2018
Received in revised form
27 August 2018
Accepted 5 September 2018
Available online 12 September 2018

ABSTRACT

This article presents the experimental data supporting analysis of differential gene expression of human cutaneous T cell lymphoma (CTCL) cell culture cells (Hut78) treated with bexarotene or a variety of rexinoids, in conjunction with "A Novel Gene Expression Analytics-based Approach to Structure Aided Design of Rexinoids for Development as Next-Generation Cancer Therapeutics" (Hanish et al. 2018). Data presented here include microarray gene expression analysis of a subset of genes. A novel method for analyzing gene expression in the context of a model of ligand mechanism, called the Divergence Score, is described. Analysis to identify the presence of potential retinoid response elements in putative promoter regions of the study genes is also presented.

Specifications table

Subject area	*Chemistry, Biology*
More specific subject area	*Rexinoid-Modulated Gene Expression*
Type of data	*Tables*

DOI of original article: https://doi.org/10.1016/j.steroids.2018.04.009
* Corresponding author.
E-mail address: Jennifer.Hackney.1@asu.edu (J. Hackney Price).

https://doi.org/10.1016/j.dib.2018.09.012

How data was acquired	*Microarray data was collected using a GenePix 4000B microarray scanner* *In silico promoter analysis was conducted via Pscan online tool*
Data format	*Analyzed*
Experimental factors	*Human CTCL cells were treated with Bexarotene or one of 12 Bexarotene Analogs for 24 hours prior to total RNA extraction*
Experimental features	*Gene expression changes exhibited by Hut78 Cells treated with Bexarotene* *compared to those treated with Bexarotene Analogs were studied*
Data source location	*Glendale, Arizona, USA*
Data accessibility	*Data is with this article*
Related research article	B. Hanish, J. Hackney Price, I. Kaneko, N. Ma, A. van der Vaart, C.E. Wagner, P.W. Jurutka, P.A. Marshall, A Novel Gene Expression Analytics-based Approach to Structure Aided Design of Rexinoids for Development as Next-Generation Cancer Therapeutics, Steroids **135**, 2018, 36-49, https://dx.doi.org/10.1016/j.steroids.2018.04.009. [1]

Value of the data

- Further analysis of data presented here may be used to better understand mechanisms of rexinoid action.
- Utilization of Divergence Score methodology can aid in the analysis of transcriptome data in relation to models of drug mechanisms of action.
- Data presented here may be used to further examine regulation of gene expression by RXR homo- and heterodimers.

1. Data

Data presented here provide information on the transcriptional response of human CTCL cells to the anticancer agent, bexarotene and 12 bexarotene analogs. In Supplementary Table 1, Gene Expression Analysis, we present the log base 2 transformed ratios of analog/bexarotene of the study genes in [1].

Tables 1–3 present data related to divergence scoring. Table 1 indicates the predictive model of the divergence score and the data from the microarray experiments of the divergence score genes. Divergence scoring is a method to determine how closely gene expression after analog treatment fits with a model for rexinoid mechanism in cancer treatment. Data used to derive divergence scores (DS) are shown below. Table 1 contains the complete set of data for the DS gene subset. Table 2 is the set of data that informs this model after fitting, with uninformative data dropped out. Table 3 derives the DS from the data set after absolute values from Table 2 are calculated.

Rexinoids such as bexarotene function, at least in part, through modulation of retinoid X receptor (RXR) activity. The RXR homo- and heterodimers bind to retinoid response elements (RREs) to regulate expression of target genes. In Supplementary Table 2, we present Z-scores for each potential RRE found in 500 bp putative promoters of each study gene in [1]. Study genes were then clustered based on the presence or absence of RREs in the 500 bp promoter region (Supplementary Table 2).

2. Experimental design, materials and methods

2.1. Gene expression analysis

Bexarotene or rexinoid (analog 1–12) were used to treat human CTCL cells (HuT78) at 1×10^{-7} M cultured in Roswell Park Memorial Institute (RPMI) Medium #1640 + 10% charcoal stripped Fetal Bovine

Table 1
Divergence data.

Gene Name	Model Prediction	Analog number											
		1	2	3	4	5	6	7	8	9	10	11	12
bad	up	0.6761	−0.3578	0.4728	−1.0657	0.1674	0.3064	0.4785	0.0978	−0.3004	−0.4722	−0.4073	0.2984
bag1	down	0.4548	−0.1931	−0.6476	−0.3607	−0.4246	−0.3723	0.4411	0.0175	0.5230	−0.4537	−0.8378	0.0488
bax	up	−1.1989	−2.6745	0.1987	−1.5414	−1.6257	1.2234	0.1637	0.4369	−0.1135	0.6851	1.4185	0.3867
bcl-2	down	0.1996	−0.6243	−0.1921	−0.1834	0.3425	0.1423	0.0398	−0.3560	−0.6557	−1.2034	0.0211	−0.2166
bid	up	0.6934	−0.5899	0.1468	−0.3541	0.3369	−0.2576	−0.3227	−0.2933	0.1794	−0.8572	−0.1396	0.4673
bim	up	0.9758	0.3634	0.1153	−0.3179	0.9035	−0.4478	−0.0912	−0.1992	0.2782	−0.7361	−0.1735	0.4925
birc5	down	−0.7392	−1.5652	−1.0879	−2.6640	−2.1280	−0.4487	−0.0786	−0.6959	−0.0467	−0.1497	−1.3048	0.4474
bok	up	0.9369	0.5623	−0.1994	0.2573	0.8504	1.1839	0.8759	0.3021	0.2098	0.2085	0.6081	0.7870
casp8	up	0.7232	0.1376	−0.0541	−0.3504	0.8407	0.0971	0.2659	0.1460	0.0610	0.0936	0.7901	0.2114
cflar	down	0.6045	0.3151	0.2066	−0.3446	0.3525	−0.3832	0.0206	−0.0824	0.0798	−0.3127	0.9081	0.4438
cdkn1a	up	1.6225	0.4978	0.7606	0.4310	0.6020	−0.2474	0.5801	0.5779	0.2611	−0.2392	0.5982	0.3592
mdm2	down	0.4228	0.0222	−0.3546	−0.5164	−0.1148	−0.8187	−0.1665	−0.4420	−0.5936	−1.0019	−0.5927	0.5068
puma	up	0.6312	−0.4092	−0.1549	−0.1997	0.2122	−0.6671	0.2054	−0.0324	−0.2898	−0.4899	−0.3574	−0.2934
rab5a	up	0.6929	0.5028	−0.0840	0.1572	0.4780	−1.5667	−0.7366	−0.1218	0.6528	−0.0780	−1.1086	0.4096
ran	down	−2.5321	1.9204	−1.4936	−3.4434	−3.5508	−0.8060	0.9631	1.1416	0.1318	−0.8111	1.3922	0.8544
rb1	up	0.6156	0.3267	0.0866	0.5494	0.2615	0.0040	0.2142	0.3238	−0.0160	−0.2965	0.2541	0.1380
rbl1	up	0.8799	−0.0669	0.1815	0.0196	0.5408	−0.2203	0.0925	−0.3215	−0.4591	−0.8737	−0.0646	0.0793
rbl2	up	0.8639	0.5003	−0.0811	0.1625	0.4762	−1.5662	−0.7340	−0.1206	0.6467	−0.0763	−1.1084	0.4110
rhoa	down	−0.0656	−1.9386	0.1181	0.0495	−1.5142	−1.3342	−1.0559	−0.4788	−0.9332	−0.3598	−1.1574	0.5640
tp53	down	0.5387	−0.0917	0.1000	−0.1566	0.3286	−0.3288	0.4028	0.1485	−0.0705	−0.2688	0.1822	0.3573

Expression data is shown for the subset of study genes which comprise the divergence scoring group. Gene names are on the left, followed by the divergence model prediction in the middle column with green representing a prediction of upregulation being an improvement to bexarotene and red representing a prediction of downregulation being an improvement to bexarotene. The right side columns contain fold-change differences of the given analog compared to its parent molecule, bexarotene.

Table 2
Model fitting.

Gene Name	Model Prediction	Analog number											
		1	2	3	4	5	6	7	8	9	10	11	12
Bad	up		−0.3578		−1.0657					−0.3004	−0.4722	−0.4073	
bag1	down	0.4548						0.4411	0.0175	0.5230			0.0488
Bax	up	−1.1989	−2.6745		−1.5414	−1.6257				−0.1135			
bcl-2	down	0.1996				0.3425	0.1423	0.0398				0.0211	
Bid	up		−0.5899		−0.3541		−0.2576	−0.3227	−0.2933		−0.8572	−0.1396	
Bim	up				−0.3179		−0.4478	−0.0912	−0.1992		−0.7361	−0.1735	
birc5	down												0.4474
Bok	up			−0.1994									
casp8	up			−0.0541	−0.3504								
Cflar	down	0.6045	0.3151	0.2066		0.3525		0.0206		0.0798		0.9081	0.4438
cdkn1a	up						−0.2474				−0.2392		
mdm2	down	0.4228	0.0222										0.5068
Puma	up		−0.4092	−0.1549	−0.1997		−0.6671		−0.0324	−0.2898	−0.4899	−0.3574	−0.2934
rab5a	up			−0.0840			−1.5667	−0.7366	−0.1218		−0.0780	−1.1086	
Ran	down		1.9204					0.9631	1.1416	0.1318		1.3922	0.8544
rb1	up									−0.0160	−0.2965		
rbl1	up		−0.0669				−0.2203		−0.3215	−0.4591	−0.8737	−0.0646	
rbl2	up			−0.0811			−1.5662	−0.7340	−0.1206		−0.0763	−1.1084	
Rhoa	down			0.1181	0.0495								0.5640
tp53	down	0.5387		0.1000		0.3286		0.4028	0.1485			0.1822	0.3573

Data points which agree with the divergence model predictions for improvements to bexarotene are dropped out of the data matrix. The remaining data represent expression data points (right columns) that run contrary to the model predictions (middle column). Gene names are listed in the left column.

Table 3
Divergence score calculation.

Gene Name	Model Prediction	Analog number											
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
bad	Up		0.3578		1.0657					0.3004	0.4722	0.4073	
bag1	down	0.4548						0.4411	0.0175	0.5230			0.0488
bax	up	1.1989	2.6745		1.5414	1.6257				0.1135			
bcl-2	down	0.1996				0.3425	0.1423	0.0398				0.0211	
bid	up		0.5899		0.3541		0.2576	0.3227	0.2933		0.8572	0.1396	
bim	up				0.3179		0.4478	0.0912	0.1992		0.7361	0.1735	
birc5	down												0.4474
bok	up			0.1994									
casp8	up			0.0541	0.3504								
cflar	down	0.6045	0.3151	0.2066		0.3525		0.0206		0.0798		0.9081	0.4438
cdkn1a	up						0.2474				0.2392		
mdm2	down	0.4228	0.0222										0.5068
puma	up		0.4092	0.1549	0.1997		0.6671		0.0324	0.2898	0.4899	0.3574	0.2934
rab5a	up			0.0840			1.5667	0.7366	0.1218		0.0780	1.1086	
ran	down		1.9204					0.9631	1.1416	0.1318		1.3922	0.8544
rb1	up									0.0160	0.2965		
rbl1	up		0.0669				0.2203		0.3215	0.4591	0.8737	0.0646	
rbl2	up			0.0811			1.5662	0.7340	0.1206		0.0763	1.1084	
rhoa	down			0.1181	0.0495								0.5640
tp53	down	0.5387		0.1000		0.3286		0.4028	0.1485			0.1822	0.3573
Total Divergence =		3.4193	6.3561	0.9981	3.8787	2.6493	5.1154	3.7519	2.3964	1.9134	4.1191	5.8630	3.5159
Average Divergence =		0.3256	0.3178	0.0499	0.1939	0.1325	0.2558	0.1876	0.1198	0.0957	0.2060	0.2931	0.1758

Data are transformed into absolute values, summed, and averaged, per analog, across the total number of genes from the divergence group. Gene names are listed on the left, the divergence model prediction is listed in the center column, and the absolute values of non-compliant divergence model data points are listed in the right columns. Sums of these values are listed along the bottom, along with the average divergence, representing the divergence score for each analog.

Serum (FBS) + Sodium Pyruvate (NaPyr) + Penicillin/Streptomycin (P/S). After a 24-hour treatment period, cells in media were centrifuged 15 mL conical tubes for 5 min at 300 *g*. Then 1 mL phosphate-buffered saline (PBS) was added to each tube, cells were centrifuged, and the supernatants were then aspirated, and 1 mL of cold PBS was added to each group. Cells were again harvested by centrifugation and treated with Aurum Total RNA Lysis solution (Bio-Rad, Hercules, CA). The RNA yield was quantified via UV spectrophotometry and the RNA quality was estimated using the A260/A280 and A260/A230 ratios. RNA concentrations from each treatment were in the range of 0.40 μg/μl to 0.80 μg/μl, with an average concentration of 0.60 μg/μl.

RNA from cells were thawed as a pair (bexarotene treated and analog treated) and hybridized with 1 μl of reverse transcriptase (RT) and random primed primer (1x Cy3 green and 1x Cy5 red) utilizing an Array 350 kit (Genisphere, Carlsbad, CA). The tubes containing each treatment were brought to concentration parity using nuclease-free water such that both final 11 μl volumes contained an RNA concentration of 0.2 μg/μl. The tubes were heated, subsequently placed on ice, and then added to a reaction mix composed of Invitrogen SuperScript II first-strand buffer (Invitrogen, Carlsbad, CA), more nuclease-free water, dNTP mix, an RNase inhibitor, and RT enzyme.

The cDNA synthesis reactions (Cy3 and Cy5) were incubated and then had their reactions halted with the addition of solution containing 0.5 M sodium hydroxide (NaOH) and 50 mM ethylene-diaminetetraacetic acid (EDTA). The tube contents were then again incubated and the reaction was neutralized with 1 M Tris–HCl to a pH of 7.5. The neutralized solutions were combined and a Qiagen PCR cleanup kit (Qiagen, Germantown, MD) was employed to isolate the cDNA portion. Vacufugation was performed to bring the labelled cDNA to a 10 μl. Concentrated cDNA was combined with 2 μl of locked nucleic acid (LNA) dT blocker, 17 μl of nuclease-free water, and 29 μl of a 2X formamide-based hybridization buffer.

Human MI HEEBO ReadyArrays from Microarrays Inc. (Huntsville, AL) were prehybridized and blocked using a bovine serum albumen (BSA) ssDNA solution to reduce non-specific binding events. Slides were incubated and washed by gentle rocking in vials with 3 M NaCl/0.3 M sodium citrate (20xSSC), 10% sodium dodecyl sulfate (SDS), and BSA solution. Each slide was then transferred to a series of separate vials and washed five times by gentle agitation in millipore water. Rinsed and dried slides had the cDNA solution applied to them while warmed. Each was covered, stored in an individual humidified hybridization chamber, and placed in a hybridization oven with gentle agitation.

After incubation, the slides were agitated in a shaker with a series of progressively less concentrated treatments of SSC and SDS. After their final wash, slides were dried and treated for visualization with a solution composed of nuclease-free water, 2X formamide-based hybridization buffer, Cy3 and Cy5 3DNA capture reagents. Slides were then once again incubated in a humidified hybridization chamber. Post-DNA hybridization washes were performed similar to pre-hybridization washes, and then the chips were dried and scanned by a GenePix 4000B microarray scanner using Genepix Pro® 7 Microarray Acquisition and Analysis software. Data is presented as log base 2 transformed analog/bexarotene ratio.

2.2. Divergence score calculation

Divergence scoring (DS) was used in [1] to further differentiate between analogs. Twenty well-characterized genes were selected as a sub-set of the study group (Table 1). These genes were chosen using two criteria, their ability to differentiate the analogs from each other and the significance of directionality of regulation. Divergence in this case means the extent to which the gene expression of each individual gene elicited by the analog was performing contrary to the hypothesized direction (increased rather than decreased for example) of expression of that gene based on its ontology.

For every DS gene from each analog, fold-change differences from the parent molecule were either discarded or aggregated depending on whether the gene expression ratio matched the model for rexinoid function (Table 1). The predictive model operates on a fundamental design principle that the analog should be an improved version of Bexarotene. When the gene expression data did fit the predicative model, the difference between the analog and Bexarotene was discarded (Table 2). For the instances in which there was a mismatch between the directional expression difference of the analog and bexarotene, and that mismatch did not align with the predictive model, the absolute value of the observed data point was recorded and then aggregated (Table 3). The aggregate value of recorded mismatches was averaged across the total number of DS genes, and the DS for the given analog was recorded as this value (Table 2).

2.3. Promoter analysis

Gene names were converted to National Center for Biotechnology Information (NCBI) Reference Sequence Database ID numbers (RefSeq ID) using the ID conversion tool available from DAVID (Database for Annotation, Visualization and Integrated Discovery) [2,3]. A 500 base pair (bp) presumptive promoter region located at −450 bp through +50 bp relative to the transcription start site was analyzed for each study gene [4]. The 500 bp regions were scanned for the presence of RXR homo- and heterodimer binding motifs as defined using well-defined Position Frequency Matrices (PFMs) from the JASPAR 2016 database (Supplementary Table 2) [5,6].

Z-scores were determined for each binding motif for each of the study genes (Supplementary Table 2). Positive Z-scores suggested that the promoter region in question was more likely to contain the binding motif when compared to the background whole genome promoter set. The resulting Z-scores were used to calculate computationally-determined matching scores. Genes were determined to contain the motif if the matching score was higher than expected when compared to the whole genome promoter set. Study genes were analyzed based on the absence or presence of each of the binding motifs using ClustVis and hierarchical cluster analysis based on Euclidian correlation distances and complete linkage methods [7].

Acknowledgements

This work was supported by the National Institutes of Health National Cancer Institute (Grant 1 R15 CA139364-01A2) and a Women and Philanthropy Grant (ASU Foundation LTR 3/16/15), both to CEW, PWJ, and PAM. Thanks to Julie Furmick for initial microarray experimental work. PAM wishes to thank the Genome Consortium for Active Teaching for initial microarray training and support for this project. We thank AzTE for ongoing intellectual property support of our work.

Transparency document. Supporting information

Transparency data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.dib.2018.09.012.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.dib.2018.09.012.

References

[1] B. Hanish, J. Hackney Price, I. Kaneko, N. Ma, A. van der Vaart, C.E. Wagner, P.W. Jurutka, P.A. Marshall, A novel gene expression analytics-based approach to structure aided design of rexinoids for development as next-generation cancer therapeutics, Steroids 135 (2018) 36–49. https://doi.org/10.1016/j.steroids.2018.04.009.
[2] D.W. Huang, B.T. Sherman, R.A. Lempicki, Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources, Nat. Protoc. 4 (2009) 44–57. https://doi.org/10.1038/nprot.2008.211.
[3] D.W. Huang, B.T. Sherman, R.A. Lempicki, Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists, Nucleic Acids Res 37 (2009) 1–13. https://doi.org/10.1093/nar/gkn923.
[4] F. Zambelli, G. Pesole, G. Pavesi, Pscan: finding over-represented transcription factor binding site motifs in sequences from co-regulated or co-expressed genes, Nucleic Acids Res 37 (2009) W247–W252. https://doi.org/10.1093/nar/gkp464.
[5] A. Mathelier, O. Fornes, D.J. Arenillas, C. Chen, G. Denay, J. Lee, W. Shi, C. Shyr, G. Tan, R. Worsley-Hunt, A.W. Zhang, F. Parcy, B. Lenhard, A. Sandelin, W.W. Wasserman, JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles, Nucleic Acids Res 44 (2016) D110–D115. https://doi.org/10.1093/nar/gkv1176.
[6] A. Sandelin, W. Alkema, P. Engström, W.W. Wasserman, B. Lenhard, JASPAR: an open-access database for eukaryotic transcription factor binding profiles, Nucleic Acids Res 32 (2004) 91D–94D. https://doi.org/10.1093/nar/gkh012.
[7] T. Metsalu, J. Vilo, ClustVis: a web tool for visualizing clustering of multivariate data using Principal Component Analysis and heatmap, Nucleic Acids Res 43 (2015) W566–W570. https://doi.org/10.1093/nar/gkv468.