Cell-Specific Computational Modeling of the PIM pathway in Acute Myeloid Leukemia

**Supplemental Materials and Methods**

Dana Silverbush1,2^, Shaun Grosskurth3^, Dennis Wang4, Francoise Powell3, Bertie Gottgens5, Jonathan Dry3\* and Jasmin Fisher2,6\*

1Department of Computer Science, Tel-Aviv University, Tel-Aviv, Israel;

2Microsoft Research, Cambridge, UK;

3AstraZeneca Oncology IMED, Waltham MA;

4AstraZeneca Oncology IMED, Cambridge UK;

5Department of Haematology, Cambridge Institute for Medical Research & Wellcome Trust and MRC Stem Cell Institute, University of Cambridge, UK;

6Department of Biochemistry, University of Cambridge, UK

^ These authors contributed equally to this work

\*To whom correspondence should be addressed:

jf416@cam.ac.uk and jonathan.dry@astrazeneca.com

**Cell line treatment**

Cell lines (CMK, EOL1, HL60, KASUMI3, KG1A, MOLM13, MOLM16, MONOMAC6, MV411, NOMO1, OCIAML2, OCIM1, and OCIM2) were purchased from American Type Culture Collection (ATCC,[*http://www.atcc.org/*](http://www.atcc.org/)) cell bank and passaged in our laboratory for fewer than 6 months after receipt or resuscitation.  ATCC uses morphology, karyotyping, and PCR based approaches to confirm the identity of human cell lines and to rule out both intra- and interspecies contamination.

All cells were cultured and assayed as previously described [1] and in supplemental methods. Briefly, cells were cultured in RPMI 1640 media (Invitrogen) supplemented with 10% FBS (Hyclone). Cells were maintained at 37oC in 5% CO2. For single agent and combination compound treatment, cells were plated onto a 96 well plate at a density of 20,000 cells per well. Cells were treated with dose response of compound over a log third scale starting at ranges from 0.003uM to 10uM based around the IC50 value for each agent. All doses were run in triplicates. For AZD1208 resistant cell lines, 4 separate pools of MOLM16 cells were induced resistant by successive doubling growth in GI25 AZD1208 over a four month period.

**Gene expression microarray**

Cell line lysate was generated from logarithmic growing CMK, EOL1, HL60, KASUMI3, KG1A, MOLM13, MOLM16, MONOMAC6, MV411, NOMO1, OCIAML2, OCIM1, and OCIM2 cell lines. Lysate was sent to Expression Analysis (http://www.expressionanalysis.com/) for gene expression analysis on Affymetrix Human Genome U133 Plus 2.0 Array. Expression results contained in CEL files were submitted to NCBI GEO under series GSE57083. The CEL files were fRMA normalized with the frma package [2], log2 transformed, and expression was averaged by gene symbol across probesets.

**Whole exome DNA sequencing**

Cell lines lysate was generated from logarithmic growing CMK, EOL1, HL60, KASUMI3, KG1A, MOLM13, MOLM16, MONOMAC6, MV411, NOMO1, OCIAML2, OCIM1, and OCIM2. Lysate was sent to Expression Analysis (http://www.expressionanalysis.com/) for whole exome DNA sequencing. FASTQ files were trimmed and aligned with the BCBio pipeline (<https://bcbio-nextgen.readthedocs.org>). Paired analysis of the parental and resistant cell lines was performed to using FreeBayes [3], MuTect [4], and VarDict (GitHub) to call resistance specific mutations.

**Theranostics health reverse phase protein array**

Cells were treated with AZD1208 or ACC220 as single agent or in combination for 3 or 24 hours. Lysate was prepared and shipped to Theranostics Health (Rockville, Maryland) for reverse phased protein array experiments. In brief detail, cell lysate samples are printed at both 0.25 and 0.5 mg/ul in quadruplicates on a nitrocellulose-coated microarray slide with an Aushon A2470 solid pin robotic printer. Spotting also includes presence of positive controls and calibration standards. Replicate arrays are spotted with identical content and stained with SYPRO Ruby Protein blot stain (Invitrogen Cat. No. S-11791) to measure protein spot amount. Each array is then stained with a unique primary fluorescently-labeled antibody that is scanned at a wavelength of 635nm. Scanned images are corrected for spatial and total protein bias in the raw data and quadrant median normalized (QMN). The QMN values were log2 transformed and separated into two datasets based on the two spotting concentration. Wilcoxon Rank Sum Tests were performed between various treatment groups using R (http://www.R-project.org). Values with difference of at least log2 fold-change of 0.5 and a raw p-value of 0.05 were considered statistically significant.

**Protein array data transformation for executable network model construction**

The relative linear log2 RPPA values were categorized for use in executable network modelling (Fig S4). Since the RPPA data is a matrix with a cell+treatment dimension and a measured protein antibody dimension, the data was categorized by cell+treatment to model the dynamics of change within a cell at a specific treatment and by protein antibody to model the dynamics of change for a protein across all cell+treatment conditions (Fig S4, Data set S5). Since most log2 RPPA distributions were reasonably normal and centered on similar means, categorization was performed by utilizing the distribution along one of the two dimensions and converting those values to a categorical value dependent on location in the distribution. Each log2 RPPA value was converted to a categorical value based on position in the distribution where values were converted to 0 if the value was a low outlier, 1 if the value was in the lower 25%, 2 if the value was in the 25-75% distribution centered on the mean, 3 if the value was in the upper 75%, and 4 if the value was a high outlier. Then, the categorized values were collapsed on the mode of all the replicate cell+treatment spots for a specific protein antibody, reducing the data to one value for each cell+treatment and protein antibody combination. Finally, a categorical value was also provided for cell growth and apoptosis at each cell+treatment condition using growth characteristics and western data (Data set S7).

**PhosphoScan phosphorylation proteomics**

MOLM16 cells were grown in RPMI media supplemented with  10% FBS and 2mM  L-glutamine. Cells were split to 0.5e6 cells/mL the night before the experiment, and grown to ~ 0.7e6/mL at time of treatment with either 2uM AZD1208 or DMSO control for 3 hours. After treatment, cells were pelleted, stored at -70°C, and shipped to Cell Signaling (Danvers, Massachusetts) for use as samples for PhosphoScan Phosphorylation Proteomics. In brief detail, cell pellets were lysed and phospho-antibody pull down was performed with the following Cell Signaling PTMScan motif antibodies: Akt/Akt (#9614/#10001), PKA/PKD (#9624/#4381), PKC/tXR (#6967/#2351), and Akt/PI3K (PTMScan Direct). LC-MS/MS analysis using LTQ-Orbitrap-VELOS, ESI-CID was performed by Cell Signaling with each sample ran in technical duplicates. Peptide calling was performed by Cell Signaling using Seaquest 3G and Sorcerer 2 search results. Criteria for reliable detection were phosphor-peptides that had a maximum percentage CV range from 137% to 0.5% and a maximum intensity range from 750,000,000 to 6,000. Phospho-peptides that had fold-changes >2.5 or <-2.5, maximum CV range of 65% to 0.7%, and maximum range 750,000,000 to 75,000 were considered meaningful (Data set S8). Intersecting the proteins measured by both RPPA and PhosphoScan resulted in 9 overlapping proteins with 4 proteins having identical phosphosites measured (Data set S9).

**Qualitative networks (QN)**

Qualitative Networks extend Boolean networks by allowing variables to range over larger discrete domains and replacing Boolean functions by algebraic functions [5]. Intuitively, a Qualitative Network associates a discrete variable with every substance the model follows. The variable ranges over a small discrete domain where values represent expression levels of the substance such as {0=off, 1=low, 2=medium, 3=high, 4=maximum observed}. For every substance, a target function reads the values of other substances that affect it. The target function sets a value to which the substance should get to eventually. The substance changes gradually to attain this target.

More formally, a Qualitative Network is defined as follows. A Qualitative Network is $Q=(V,T,N)$ consists of a set of variables $V=\left\{v\_{1},…, v\_{n}\right\}$ ranging over $\left\{0,…,N\right\} $and a set of target functions $T=\left\{T\_{1},…,T\_{n}\right\}$. A state of the network is $s:V\rightarrow \{0,…,N\}$, i.e., a valuation for all the variables in $Q$. A target function $T\_{i}$ associates with a state $s$ a value in $\left\{0,…,N\right\}$, the value towards which $v\_{i}$ should move from state $s$.

A Qualitative Network gives rise to a transition system between its states. As explained, the values of substances/variables change by pursuing their targets gradually, i.e., they change by at most 1 in every transition. Thus, every state $s=\left\{d\_{1},…, d\_{n}\right\}$, where the value of $v\_{i}$ is $d\_{i}$, has a successor $s’=T(s)$, where $s^{'}\left(v\_{i}\right)$ is defined as follows:

$$\begin{matrix}d\_{i}-1&if&T\_{i}\left(s\right)<d\_{i} and d\_{i}>0 \\d\_{i}+1&if&T\_{i}\left(s\right)>d\_{i} and d\_{i}<N \\d\_{i}&if&\left\{\begin{matrix}1. T\_{i}\left(s\right)=d\_{i} \\2. T\_{i}\left(s\right)<d\_{i} and d\_{i}=0 \\3. T\_{i}\left(s\right)>d\_{i} and d\_{i}=N\end{matrix}\right. \end{matrix}$$

Notice that all variables are updated simultaneously. A discrete change in the value of a variable corresponds to a number of simultaneous molecular events over a period of time, and as such non-deterministic events are not considered. Target functions default to the difference between the average of the activator nodes and the average of the inhibitor nodes unless explicitly altered.

It follows that a Qualitative Network defines a finite state transition system where state changes are deterministic. Then every execution of a Qualitative Network ends in a cycle of states that are visited infinitely often. We say that a state is *recurring* if it appears in an execution that starts from itself. That is, if for some finite number of applications of $T$, we have $s=T^{i}(s)$. A network is *stabilizing* if there is a unique state $s$ such that $s=T(s)$ and no other state is recurring.

**Construction of a cell-specific AML model**

An initial AML QN model containing 30 nodes and 64 interactions was generated from the manual curation of 68 publications (see Table S7 for full literature entries). The literature model was further refined by iterative modification of the target functions. The target function for each internal node was iteratively modified to reflect the levels of phosphorylation activity as measured by the transformed RPPA data as well as the qualitative activity reported in the literature. To optimally utilize the RPPA data, specific phosphorylation sites reported in the literature that drive signaling pathways were used as guides for model consistency. For example, the phosphorylation of 4EBP1 by mTOR occurs at multiple sites, but the phosphorylation at S65 of 4EBP1 causes the greatest decrease in affinity of 4EBP1 for eIF4E [6]. Therefore, we chose to focus on the RPPA values measured at S65 instead of capturing the total protein values of 4EBP1 to accurately focus on the in-questioned signaling cascade. We further tuned the model to the AML disease state. For example, the model reflects AML pathways in which the phosphorylated 4EBP1 triggers over-activation of EIF4E, thus 4EBP1 is modeled to activate EIF4E as opposed to its healthy state in which unphosphorylated 4EBP1 inhibits eIF4E activity through suppression of mRNA translation [7]. Similarly, BAD is modelled to inhibit apoptosis, as its phosphorylation disrupts healthy state bindings and allows BAD to bind scaffold protein 14-3-3 to sequester its proapoptotic function, thereby activating a cell survival pathway [8], [9]. Additionally, varying magnitudes of influence by different factors were incorporated as “influence-weight” for each protein node in each interaction. For example, both AKT and PIM1 phosphorylate PRAS40 on T246 leading to its dissociation and subsequent activation of mTORC1. However, AKT had a more dominant effect on mTORC1 activity in the RPPA data, so the target function of mTORC1 gives more weight to AKT over PIM1. Nodes that had no impact on the activity and cell behavior were removed, such as FOXO3, or simplified to a complex, such as collapsing the nodes for Grb2 and SOS.

**References**

[1] E. K. Keeton et al., “AZD1208, a potent and selective pan-Pim kinase inhibitor, demonstrates efficacy in preclinical models of acute myeloid leukemia,” Blood, vol. 123, no. 6, pp. 905–913, Feb. 2014.

[2] M. N. McCall, B. M. Bolstad, and R. A. Irizarry, “Frozen robust multiarray analysis (fRMA),” Biostat. Oxf. Engl., vol. 11, no. 2, pp. 242–253, Apr. 2010.

[3] E. Garrison and G. Marth, “Haplotype-based variant detection from short-read sequencing,” ArXiv12073907 Q-Bio, Jul. 2012.

[4] K. Cibulskis et al., “Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples,” Nat. Biotechnol., vol. 31, no. 3, pp. 213–219, Mar. 2013.

[5] M. A. Schaub, T. A. Henzinger, and J. Fisher, “Qualitative networks: a symbolic approach to analyze biological signaling networks,” BMC Syst. Biol., vol. 1, p. 4, 2007.

[6] M. E. Feldman et al., “Active-Site Inhibitors of mTOR Target Rapamycin-Resistant Outputs of mTORC1 and mTORC2,” PLOS Biol, vol. 7, no. 2, p. e1000038, Feb. 2009.

[7] M. C. Mendoza, E. E. Er, and J. Blenis, “The Ras-ERK and PI3K-mTOR pathways: cross-talk and compensation,” Trends Biochem. Sci., vol. 36, no. 6, pp. 320–328, Jun. 2011.

[8] D. Shahbazian, A. Parsyan, E. Petroulakis, J. Hershey, and N. Sonenberg, “eIF4B controls survival and proliferation and is regulated by proto-oncogenic signaling pathways,” Cell Cycle Georget. Tex, vol. 9, no. 20, pp. 4106–4109, Oct. 2010.

[9] Y. Tan, M. R. Demeter, H. Ruan, and M. J. Comb, “BAD Ser-155 phosphorylation regulates BAD/Bcl-XL interaction and cell survival,” J. Biol. Chem., vol. 275, no. 33, pp. 25865–25869, Aug. 2000.