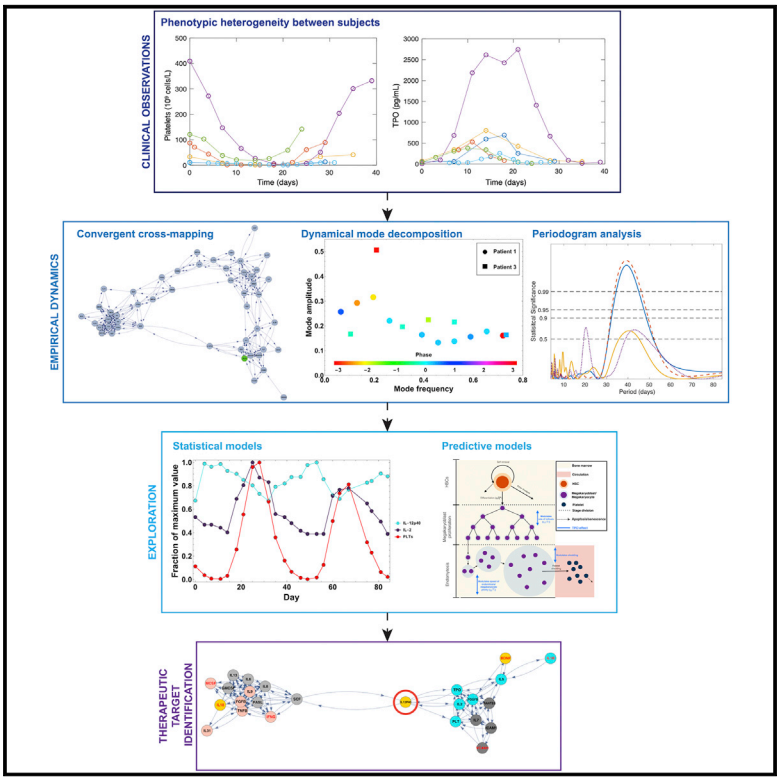
Article

Patterns

A Blueprint for Identifying Phenotypes and Drug Targets in Complex Disorders with Empirical Dynamics

Graphical Abstract



Highlights

- Many disorders are "complex" and include multiple dysfunctional elements
- Empirical dynamics can infer networks of interactions from longitudinal data
- Subnetworks around focal points of interest can explain phenotypes
- Dimensional reduction narrows the search for therapeutic interventions

Authors

Madison S. Krieger, Joshua M. Moreau, Haiyu Zhang, May Chien, James L. Zehnder, Morgan Craig

Correspondence

morgan.craig@umontreal.ca

In Brief

Many disorders are characterized by more than one aberrant biomarker. When the number of such biomarkers increases, it becomes difficult to define meaningful variation within patients as well as to identify targets for therapeutics. With such complexity and biological uncertainty, traditional techniques are ineffective. We suggest an empirical-dynamical toolbox for inferring networks of biomarker interactions. We infer and validate networks in three subjects with cyclic thrombocytopenia, confirming clinical suspicions of a novel phenotype and identifying a new therapeutic axis.

Krieger et al., 2020, Patterns *1*, 100138
December 11, 2020
https://doi.org/10.1016/j.patter.2020.100138

CellPress

Patterns

CellPress
OPEN ACCESS

Article

A Blueprint for Identifying Phenotypes and Drug Targets in Complex Disorders with Empirical Dynamics

Madison S. Krieger,[1] Joshua M. Moreau,[2] Haiyu Zhang,[3] May Chien,[4] James L. Zehnder,[3,4] and Morgan Craig[5,6,7,*]
[1]Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA, USA
[2]Department of Dermatology, University of California, San Francisco, CA, USA
[3]Department of Pathology, Stanford School of Medicine, Stanford, CA, USA
[4]Department of Medicine (Hematology), Stanford, CA, USA
[5]Département de Mathématiques et de Statistique, Université de Montréal, Montréal, QC, Canada
[6]CHU Sainte-Justine Research Centre, 3175 Chemin de la Côte-Sainte-Catherine, Montréal, QC H3T 1C5, Canada
[7]Lead Contact
*Correspondence: morgan.craig@umontreal.ca
https://doi.org/10.1016/j.patter.2020.100138

THE BIGGER PICTURE While some diseases have known root causes, many consist of multiple abnormal behaviors connected by unknown biology. These complex disorders require new tools that are able to leverage data to sketch out the equivalent of a "food web" in ecology, i.e., a network of interactions between various cells and protein signals.

Constructing these networks allows us to examine differences between healthy and afflicted individuals and identify potential therapies. Discerning meaningful differences in interaction networks also gives a paradigm for delineating phenotypes. Here, we integrate recent techniques for empirical network inference and apply them to cyclic thrombocytopenia, a complex blood disease characterized by platelet oscillations. The agreement between the techniques builds confidence in the networks we have inferred. We validated a plausible therapeutic intervention *in silico*, suggesting that interactions along the axis we identify as the problem may be clinically viable.

Proof-of-Concept: Data science output has been formulated, implemented, and tested for one domain/problem

SUMMARY

A central challenge in medicine is translating from observational understanding to mechanistic understanding, where some observations are recognized as causes for the others. This can lead not only to new treatments and understanding, but also to recognition of novel phenotypes. Here, we apply a collection of mathematical techniques (*empirical dynamics*), which infer mechanistic networks in a model-free manner from longitudinal data, to hematopoiesis. Our study consists of three subjects with markers for cyclic thrombocytopenia, in which multiple cells and proteins undergo abnormal oscillations. One subject has atypical markers and may represent a rare phenotype. Our analyses support this contention, and also lend new evidence to a theory for the cause of this disorder. Simulations of an intervention yield encouraging results, even when applied to patient data outside our three subjects. These successes suggest that this blueprint has broader applicability in understanding and treating complex disorders.

INTRODUCTION

The desirable scenario in which a single biomarker can be mapped to a single disease, which presents identically in every patient, is overly idealistic. In fact, few diseases are diagnosable via a single metric in a binary "yes/no" fashion. Even single biomarkers that are extremely indicative of particular disorders, such as for monogenic diseases, do not ensure an accurate forecasting of each patient's phenotypic presentation. Classic examples where multiple phenotypes exist for single-hit diseases

include holoprosencephaly,[1,2] epidermolysis bullosa,[3] and van der Woude syndrome,[4] among many others. In addition to having multiple phenotypes, many diseases are classified according to more than one aberrant biomarker. For instance, there is growing evidence that there may be dozens of phenotypes contained under the "umbrella" of multiple sclerosis, which are differentiated not only based on life experience and life expectancy but also based on the very biomarkers that are abnormal.[5,6] The growing recognition that many (if not most) diseases are characterized by different phenotypes has led not only to an increasing emphasis on personalized medicine, but also to the idea that the existence of multiple phenotypes can be beneficial in translating clinical observations of many phenotypes into mechanistic understanding of what unites and divides those phenotypes.[7,8] Intersecting mechanisms between different phenotypes bring us closer to describing the core etiology of the umbrella "disease" and can translate into better clinical treatment.

When there are multiple irregular biomarkers which may also belong to multiple phenotypes, the technical challenge is in identifying these phenotypes, assigning patients to phenotypes, and mapping between clinical observations and phenotypic mechanisms that explain them. To overcome this challenge, here we argue for the use of a growing body of analytical literature called *empirical dynamics*,[9,10] which we use to refer to any equation-free and non-parametric tool that seeks to infer causal mechanisms in a system or structural properties of a system from empirical data. These are well suited to our purpose because they assume no model, which fits our lack of knowledge of the biology connecting the various perturbed components, and also because the inferred mechanisms order the perturbed components into networks of cause-and-effect where no such order existed before. Our "blueprint" for bringing these techniques to bear on clinical decisions consists of (1) inferring and vetting these mechanistic networks, (2) narrowing our attention to small subnetworks surrounding clinically relevant components, and (3) modeling clinical interventions based on the subnetwork. Analyzing these networks also enhances our understanding of the underlying biology beyond a single intervention, because a component of interest can now be modeled and understood in terms of subnetworks which consist of vastly smaller numbers of components compared with the total dataset. Our battery of tests include three very different techniques from this area: convergent cross-mapping (CCM),[11–14] transfer entropy,[15] and dynamical mode decomposition (DMD).[16–19] Each of these relies on different mathematical analyses of time series data to determine whether or not two elements of a given system are mechanistically connected: CCM draws on dynamical systems theory and seeks to reconstruct local patches of a putative underlying attractor, determining that there is a mechanistic connection between two variables if they reliably reconstruct the same patches of one shared attractor; transfer entropy draws on information theory and seeks to quantify the amount of entropy that is shared between two causally linked time series; determining that there is a mechanistic connection between two variables if the mutual information clears a threshold of significance; DMD draws on linear algebra and operator theory, and seeks to enumerate a much smaller number of behaviors (modes) of a system in comparison with the number of entities that constitute that system, which, while not explicitly identifying mechanistic connections between pairs of variables, still imputes order and interaction to larger ensembles of variables. While there may well exist deeper connections to each of these techniques, such connections are far from understood, and therefore we believe consensus between them can be considered increasingly indicative of a reliable connection that can be exploited in the lab or clinic. Our selection of techniques is also commensurate with the growing wisdom that the most accurate inference of biological networks arises from consensus between orthogonal techniques, rather than from a single approach.[20,21]

As a first step to demonstrate the natural pairing and viability of these techniques to complex disorders, we applied them to a rare blood disease called cyclic thrombocytopenia (CTP).[22–24] This disease is characterized by regular oscillations in platelets as well as the cytokine thrombopoietin (TPO) which regulates the production of platelets and platelet precursors (megakaryoblasts and megakaryocytes). In addition, many patients show regular oscillations in the counts of many important signaling proteins (cytokines). The rich time behavior of this disorder has invited comparisons with canonical examples in dynamical systems and has fueled many theoretical studies.[24–28] However, the same dynamical richness that has made these systems of interest to theorists also poses challenges to experimentalists and clinicians, who rely on a particular set of tools to identify therapeutic targets and generate functional hypotheses on the etiology of a particular disorder in a particular patient. For instance, correlation analyses are easily confounded by these diseases, as entities that co-cycle with identical periods will always be found to be strongly correlated, shedding no light on the mechanistic interactions driving the oscillations. The temporal behavior of this disorder therefore presents an ideal case study in which the analytical tools we are promoting might be able to untangle mechanistic networks from data that are not amenable to correlative-type approaches.

CTP is also of immediate clinical relevancy due to difficulty of treatment,[29] especially compared with other blood disorders in which biomarkers are oscillating (such as cyclic neutropenia).[30] Despite frequently being associated with disrupted TPO receptor interactions,[24,31,32] regular treatment with exogenous TPO (the principal protein regulating the megakaryocyte/platelet axis) mimetics is generally ineffective.[29] This suggests that CTP may be mechanistically distinct from the other diseases, which are more easily treated with mimetics for the defective protein.

To demonstrate the feasibility and usefulness of empirical dynamics for our case study of CTP, we analyzed data on platelets, TPO, and more than 60 cytokines measured in time series for three individuals: subject A, who carries a heterozygous germline mutation resulting in a loss of *c-mpl* function,[33] his father, subject B; and an unrelated individual with CTP, subject C. The father-son relationship between subjects A and B is particularly interesting because, despite sharing the *c-mpl* mutation, subject B has never shown symptoms of CTP. This makes subject B a useful healthy (non-CTP) control and evidence of a possible pre-disease state. Empirical dynamics agreed on couplings uncovered in all three subjects with a remarkable degree of consensus, further corroborated by a manual verification of more than 1,000 interactions in the experimental literature. The networks we uncover are extremely different between subjects

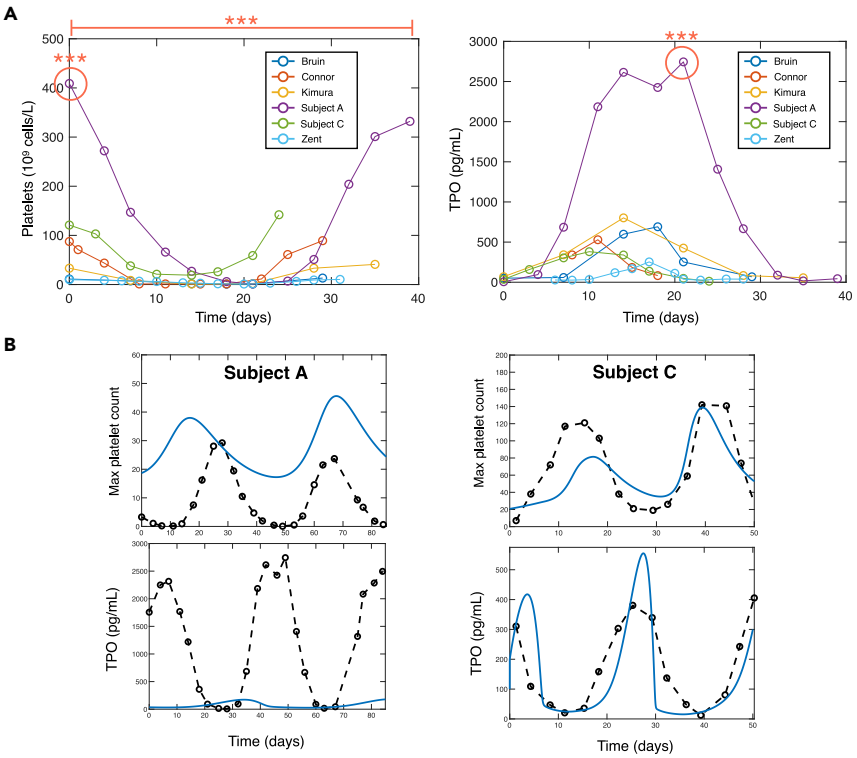


Figure 1. Phenotypic Heterogeneity in Clinical Presentation of Cyclical Thrombocytopenia Is Pronounced in Subject A

(A) Platelet and TPO concentrations over one cycle from subjects A and C, in addition to four previously characterized subjects (see Langlois et al.[24]). Asterisks indicate the number of standard deviations from the mean for maximum platelet concentration (left panel, circle), period length (left panel, line), and maximum TPO concentrations (right panel, circle) for subject A compared with the five other individuals.

(B) The fits of the Langlois[24] model for platelet (top) and TPO (bottom) levels to both subjects A and C is within the same margins of error as the other subjects in the figure, while subject A greatly exceeds these.

A and C, supporting preliminary evidence that subject A might represent a novel phenotype. Despite these differences, an unexpected connection was found in that both phenotypes implicate a single biological mechanism, the Th-17 cell differentiation and reaction pathways, as critical in causing platelet oscillations. The heterogeneity between these two individuals is explained as different dysregulations of Th-17 cell maintenance, but the implication of this axis situates these two subjects in a broader ongoing narrative about the role of Th-17 cells in cycling and non-cycling thrombocytopenia.[34,35] As a first step toward actual therapeutic intervention on this axis, we simulated such an intervention on our subject C as well as other previously examined patients who we believe belong to the primary phenotype, using an existing theoretical model.[24] Despite the different observed platelet trajectories and different fits within the model case, the simulated intervention completely removed fluctuations in the platelet line for multiple subjects, which we believe represents a step toward successful combination therapies for this disorder.

RESULTS

We began by evaluating the clinical characteristics of individuals with CTP identified by our team at the Stanford Medical Center. A total of n = 3 individuals were included in the study: subject A has a previously reported novel *c-mpl* heterozygous germline mutation affecting TPO receptor function,[33] subject B carries the same mutation but shows no clinical presentation of the disease, and subject C has CTP. Subject A visited the Stanford Medical Center clinic every 3–4 days over a period of 84 days and complete blood counts were taken; subject C visited the clinic every 3–4 days for a total of 49 days.

Preliminary Evidence for Multiple Phenotypes

Subject C has a more classical clinical presentation of CTP, with an oscillation period of roughly 27 days, similar to previously reported cases (typically between 25 and 35 days).[23,31,32,36] Subject A, despite possessing a germline mutation, only recently presented to the clinic. In this individual, we measured a much longer cycling period than other CTP cases with respect to both circulating platelet and TPO concentrations. We began our investigations by comparing platelet and TPO concentration time series over one cycle from a previous cohort[24] of individuals with CTP with subjects A and C. It was immediately apparent that subject A exhibited both higher platelet concentrations at maximum and TPO concentrations substantially higher than any of the other subjects (Figure 1A). To investigate further, we next compared the cycling period and the maximum platelet and TPO concentrations from four previously examined[24] individuals with CTP with subjects A and C (Figure 1B). In all three metrics, subject A exhibited statistically significant differences in comparison with both subject C and the four previously examined subjects (p values of 0.013, 0.005, and 0.005, respectively, from two-sided Student's t test). This suggested that, despite identical clinical diagnoses, there is a non-clinical phenotypic heterogeneity component to CTP.

Using a mathematical model of platelet production,[24] validated for individuals with a "classical" presentation of CTP, we simultaneously fit platelets and TPO from both subjects 1 and 3 to characterize the pathophysiology of each individual's disease. Given the very high circulating TPO concentrations at peak, the model was unable to recapitulate both the platelet and TPO dynamics (Figure 1B, left panel; top and bottom). In contrast, the Langlois[24] model successfully captured the oscillatory patterns in platelet and TPO concentrations from subject C (Figure 1A). These results underline the unique phenotypic presentation of CTP in subject A, and the overall phenotypic heterogeneity within cyclic thrombocytopenic patients.

To distinguish phenotypic heterogeneity in CTP beyond the behavior of TPO and platelets, we employed ELISA and a 62-plate Luminex immunoassay to quantify TPO and plasma

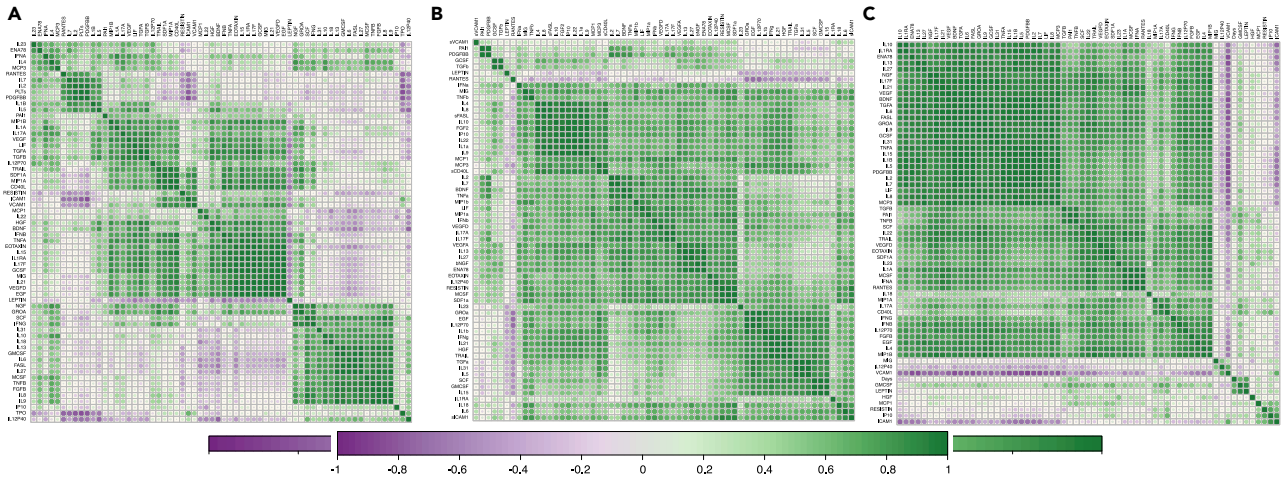


Figure 2. Structure of Hierarchically Clustered Correlations Suggests Strong Heterogeneity between Subjects A and C

Pairwise correlations between each cytokine (as well as TPO and platelets), clustered according to similarity. (A) Subject A, (B) subject B, (C) subject C. In all: purple indicates negative correlation, green indicates positive correlation. For a version without hierarchical clustering, see Figure S3.

cytokine concentrations,[33] respectively, in all three of our subjects. We have previously reported classes of cytokines cycling with identical periods in each subject through Lomb-Scargle[37,38] periodogram analysis.[24] Here, we extended our analysis by performing correlation analysis on all cytokines (as well as TPO and platelets) within each subject, including on pairs of entities that do not have identical periods. We then performed hierarchical clustering based on dissimilarity using the Lance-Williams dissimilarity update formula.

Our results show that the two symptomatic individuals (subjects A and C) have very different structure in their correlation hierarchies (see Figures 2A–2C for a version without hierarchical clustering, such that the cytokines are in the same order for every subject; see also Figure S3). Whereas subject C has most cytokines correlated and only two adhesion factors (intercellular adhesion molecule 1 [ICAM1] and vascular cell adhesion protein 1 [VCAM1]) anti-correlating, the correlation network in subject A's case has more complex structure, with several correlated clusters as well as many more cytokine pairs that are un- or anti-correlated. Despite the genetic relationship between subjects A and B and the disease-specific relationship between subjects A and C, it is actually subjects B and C who have the most similar structure in these hierarchies, suggesting that there is significant phenotypic heterogeneity between subjects A and C.

Network Inference: Identifying Distinct Coupling Networks between the Two Proposed Phenotypes

To characterize regulators of observed cytokine oscillations in these individuals, we sought to distinguish causal interactions from the correlations we identified. For this, we employed CCM, a causal network inference technique,[11–14] to provide mechanistic and etiological insight into the hematopoietic process in all three subjects (see Experimental Procedures), which we later validate using both orthogonal inference techniques and manual verification in the literature.

CCM is a particularly apt choice for this dataset, because it excels at mechanistic inference in time series with any kind of generalized recurrent or oscillatory behavior. This is due to its particularly geometric nature. The method consists of a leave-one-out cross-validation test for Takens' theorem,[39] an observation that mechanistically coupled variables have the unusual property that data points clustered in a delay embedding (a plot of the points on the axes $\{X(t), X(t-1), X(t-1), \ldots X(t-E-1)\}$ of one variable will also be clustered in the other variable. An example with a 2D ($E = 2$) embedding of TPO and platelets for subject A is shown in Figure 3A. CCM assumes the clustering described in Takens' theorem, and uses it to reconstruct a prediction for one variable using local clusters from another. If this prediction is sufficiently good, it is taken as evidence that Takens' theorem holds for that pair of variables, meaning that they are mechanistically connected (see Experimental Procedures for a more detailed algorithm, and Supplemental Experimental Procedures for the mathematical foundations). The benefit of generally recurring behaviors in the underlying time series is seen by understanding the delay embeddings as a kind of *pattern space*. For instance, in a 2D embedding, nearby points have very similar values of $[X(t), X(t-1)]$, which in the original time series are just similar line segments; likewise, in a 3D embedding, nearby points look like similar quadratic patterns in the time series (and cubic patterns in 4D, etc.). Any time series which repeats similar patterns will provide a denser resolution of points (patterns) in the delay embedding if measured often enough, which will in turn give tighter predictions in CCM and therefore sharper results when variables are actually mechanically coupled. We believe this generalized near-repetition of patterns occurs broadly in biomedical data. Even in diseased states without obvious repetition, fluctuations about a mean value still occur on time scales that are not directly tied to the disease, such as circadian rhythms, and these fluctuations still contain information about mechanistic couplings.

The networks inferred via this method for subjects A, B, and C are shown in Figures S4, 3B, S5, and S6, respectively. As in the previous correlation results, there is a substantial difference between the inferred networks for subjects A and C. Subject C has a

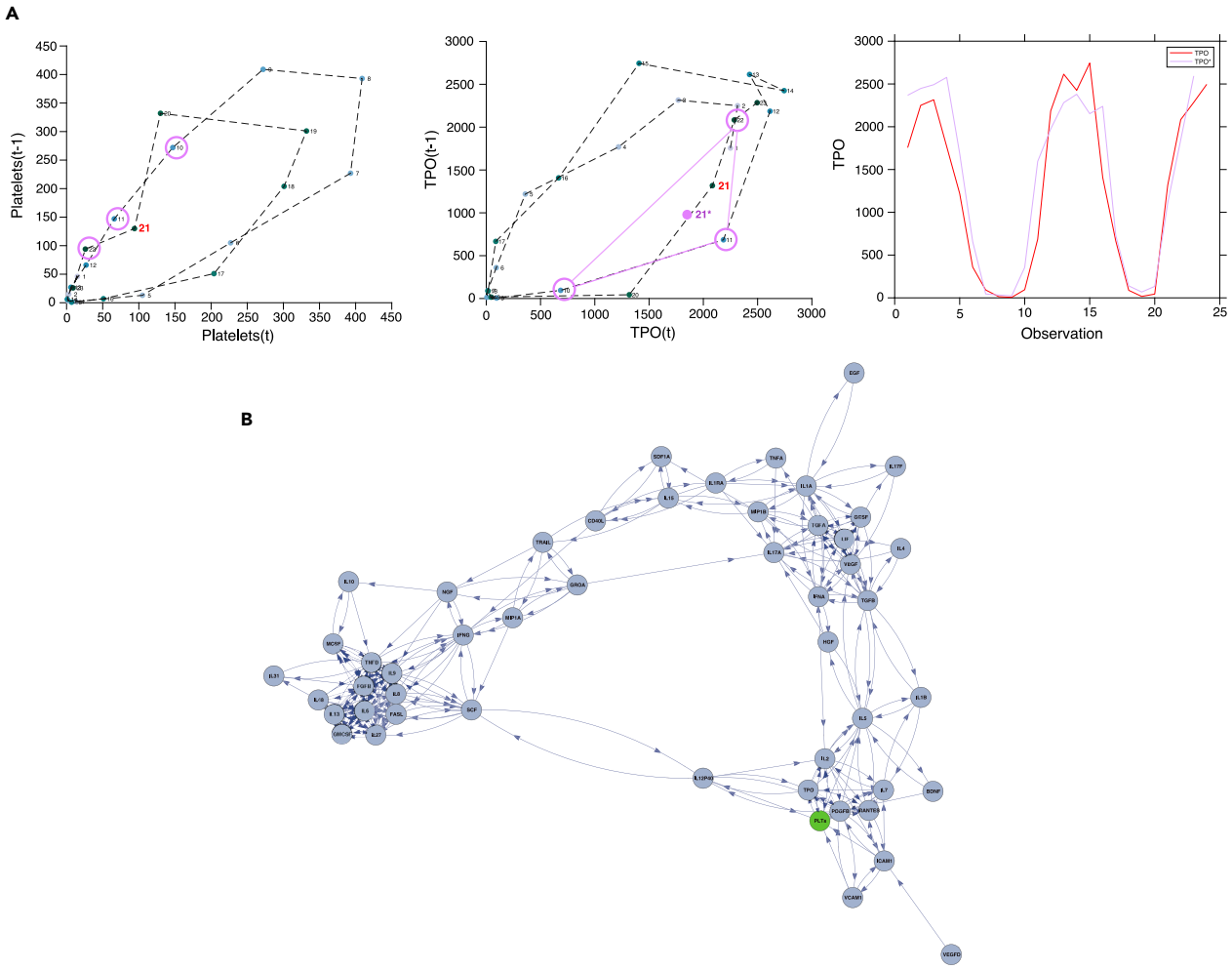


Figure 3. Mechanistic Relationships between Cytokines Are Revealed by CCM

(A) CCM infers mechanistic interactions by testing if information about the *delay embedding* (see the Experimental Procedures) of one variable is contained in another variable. Numbers (and hue) denote the observation corresponding to each point, and subsequent observations are connected by dashed lines to help visualize the geometry of the embedding. Here, we have highlighted observation 21 in the 2D delay embedding of platelets in subject A, as well as the three (because three points are necessary to triangulate a point in 2D) nearest neighbor observations (left panel). To test if platelets and TPO are interacting, we use the same three nearest neighbors to interpolate a prediction of the value of TPO in subject A at observation 21 (center panel). Repeating for every observation yields a prediction of TPO in subject A, based off the behavior of platelets (right panel). If the prediction and the true data are sufficiently close (see the Experimental Procedures), then we conclude that platelets are a mechanistic driver of TPO.

(B) Repeating the process detailed in (A) for every possible pair yields a network of interactions. Here, we show the inferred network for subject A, with platelets (the primary variable of interest in CTP) highlighted in green to aid the eye. For a comparison between the adjacency matrix and the correlation results in Figure 2, see Figure S1. Enlarged images of the networks for each subject are provided in Figures S4–S6.

well-connected network with very little structure. The interactions in subject A form a complex structure consisting of three hubs connected via bottlenecks of one to three cytokines (Figure 3). Further, these bottlenecks are of unique specificity to subject A; notably, an interaction between TPO, platelets, platelet-derived growth factor-BB (PDGFBB), and the bottleneck cytokine interleukin-12 subunit p40 (IL12-p40) was one of the strongest inferred for subject A, but is not known in the literature and was categorically rejected as a possible mechanism in his father, subject B (CCM end-library score of precisely 0 for all of these edges; see Table S1). This begins to suggest a unique role for IL12-p40 in subject A and also designates it as a marker of interest for understanding the biological mechanism of subject A's condition.

Network Validation

To ascertain the accuracy of the hematopoietic network inferred by CCM, we checked the literature for previous studies on the interactions described by each of the edges in the interaction networks inferred for subjects A and C. In both subjects, most of the inferred edges are described by previous studies (subject A: 78% (215/275) of inferred couplings supported by the literature; subject C: 76.8% (586/763) of inferred couplings supported by the literature; Supplemental Information [Spreadsheet]). The remaining 20% of the inferred interactions could therefore represent false positives, novel interactions in normal hematopoiesis not known in the literature, or novel interactions peculiar to CTP. To try to unravel which unsubstantiated interactions might

be false positives and could potentially represent true discoveries, we re-analyzed the unsubstantiated interactions found by CCM both using transfer entropy. Transfer entropy[40,41] infers mechanistic couplings in variables when they share mutual information, whereas CCM infers mechanistic couplings in variables when one variable can recapitulate information about a dynamical attractor observed in the other (Experimental Procedures and Supplemental Experimental Procedures). While it is possible that these two techniques are not completely orthogonal to one another, they rely on very different mathematical tools to perform their inference and therefore we consider positive results from both to be a stronger result than either one independently. Transfer entropy agreed with 100% (177/177) of the unsubstantiated couplings found by CCM in subject C and 25% (15/60) of the unsubstantiated couplings in subject A (see Experimental Procedures and Supplemental Experimental Procedures).

A possible source of false positives in both the correlation and CCM analysis is the fact that many cytokines cycle with exactly the same dominant period, i.e., their periodograms have maxima at the same location (see, for instance, the period classes of Figures 5 and S12). A visual intuition for why this could generate false positives can be gained from Figure 3A, which shows the delay embedding described above for both platelets and TPO, which always have the same dominant period in CTP. These two representations look rather similar in shape because the two underlying time series have similar periodograms and are measured at the same rate. Such co-cycling behavior represents a challenge to any causal inference tool, and not just CCM; for instance, if two time series are represented by a single oscillatory mode, they will always be found to be mechanistically driving one another. We have devised a null hypothesis test specifically to address this problem: if the inferred coupling is simply due to the two variables having similar periodic behavior, such a coupling should be inferred just as easily with synthetic data that have the same periodogram as the original data. If the inferred coupling is due to something more than the periodogram similarity between the two time series, than the original data should drastically outperform such synthetic data. Therefore, to investigate the remaining 75% (45/60) of the unsubstantiated edges in subject A that were not also inferred by transfer entropy, we generated 100 surrogate time series for the variables in question, using a technique that ensures each surrogate has the same periodogram as the time series of the variable it is meant to replace.[40] An example of such a surrogate is illustrated in Figure S15. If the original variable had a CCM score (see Experimental Procedures) in the top 5% of all of the surrogates with the same periodic behavior, we considered that as possible additional evidence that these were not false positives, or that the source of the false positive was not the periodic behavior. Repeating the CCM analysis with the surrogate time series returned positive couplings on 67% of the unsubstantiated couplings from the first analysis, including each of the 25% of the unsubstantiated couplings that were found as positives by transfer entropy. Of the unsubstantiated couplings, 20/60 (33%) found by CCM had no further support.

We next computed the Pearson correlation between possible cytokine-cytokine pairs for subject A and used the same confidence threshold as for CCM (Pearson's $R > 0.8$ with $p < 0.07$) to induce an undirected interaction network for comparison (Table S3; Figure S2). This network had only 58% (187/328) of its edges corroborated by the literature. To provide a null hypothesis for comparison, we drew 100 random "couplings" from the 4,032 possible pairs of cytokine interactions and again searched for any experimental validation, finding support for 37% (37/100; Table S2).

Finally, we examined the cytokine time series using DMD[16] (Experimental Procedures and Supplemental Experimental Procedures). DMD uses linear analysis tools based on operator theory to decompose all of the time series into a much smaller number of modes—collections of variables (potentially overlapping between modes), which, as an ensemble, fit well to complex exponential functions. As such, it is not designed to test every single possible pair of variables and infer whether or not they are mechanistically connected, because no information is provided below the level of which variables belong to which modes, and there can be dozens of variables in a single mode. Our analysis was intended to provide additional insight on the timescales relevant to the disease rather than to probe the various couplings between cytokines from yet another angle. However, we nonetheless found that the cytokines in each mode provided a remarkable degree of consensus on the networks of couplings in each subject. In Figure 4, for instance, we superpose the groupings of cytokines into the leading DMD modes for subject A atop the network of couplings inferred by CCM. To the eye, the three leading modes correspond highly to the three hubs in the network, and cytokines that are shared by multiple modes correspond highly to the bottleneck cytokines that connect the hubs.

Subnetwork Analysis: Identifying IL12-p40 as a Potential Novel Therapeutic Target for Subject A

Lomb-Scargle periodogram analysis on our data reveals that, in subjects A and C, many cytokines have identical periods to one another. In subject C, all cytokines (save one) fall into one of two classes: not cycling, or cycling with a period of 26.25 days (Figure S11). Subject A perfectly divides into six classes: not cycling, and cycling with periods of 29.2, 31.9, 35.1, 39.0, and 43.8 days. This further highlights the difference between the two subjects, and also offers more information to leverage understanding how the various cycling cytokines are connected. Given the degree of pairwise correlations within each class, we anticipated the mechanistic coupling networks to be complete graphs (each node connected to every other by a unique edge). The 29.2 and 31.9 day classes almost achieve completeness, save for one or two edges (in the 29.2 class, for example, interferon-γ (IFN-γ) and IL-31/macrophage colony-stimulating factor (MCSF) do not cross-map, nor do MCSF and IL-31, whereas stem cell factor (SCF) does not cross-map to IL-13 in the 31.9 class); however, interesting structures emerge in the 35.1, 39.0 (the period of oscillations in the platelets and TPO), and 43.8 classes where graphs were not complete. In the 39.0 day class, IL-1β only sends and receives information from IL-5 and functions as a controller outside of the complete subgraph formed by IL-5, IL-2, TPO, PDGFBB, and circulating platelet concentrations. Similarly, in the 43.8 class, VCAM1 sits atop the complete subgraph comprised of ICAM1, IL-7, and RANTES (regulated on activation, normal T cell expressed and secreted), and interacts solely with VCAM1. Most curiously, there are no edges between

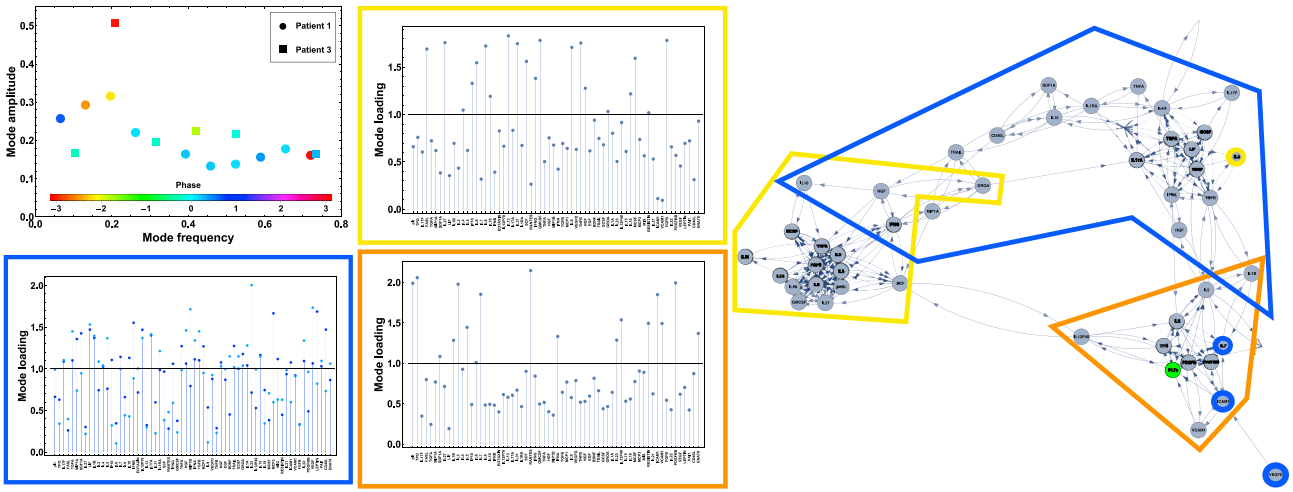


Figure 4. Dynamical Mode Decomposition Recapitulates the Network Structure and Elucidates Timescale Dynamics

Dynamical mode decomposition (DMD isolates complex exponential modes in data. Top left: the leading modes in subject A are shown according to their amplitude, period, and relative phase. For all modes, the real part of the exponential is zero (stable modes). Stem plots: the loadings for the leading four modes in subject A. The horizontal lines are merely an aid for the eye. Network: the leading three modes recapitulate the network structure found by CCM to a surprising degree, but also reveal broader coupled oscillations. For instance, most of the cytokines in the third mode are not found to be cycling by individual periodograms, but DMD suggests that they may be cycling with periods longer than any of the cytokines for which the periodogram confidence was high. Three nodes in the network (EGF, MIP1B, and BDNF) occur in none of the top four modes with very high loading and have been grayed out. Expanded figures of mode loadings are provided in Figures S8–S10.

the members of the 35.1 class (IL18, IL12-p40, and brain-derived neurotrophic factor [BDNF]). One of these (IL-12p40) appears as a critical bottleneck mitigating information from the shorter-period cytokines (29.2 and 31.9 day classes) and the longer-period cytokines (39.0 and 43.8 day classes). This is further support for the importance of this cytokine, which was the source of several of the unique (unsubstantiated by previous experiments) couplings found in the previous section. We consider the fact that this cytokine is found by all of our analyses to be directly coupled with platelets to be significant, especially as its biology is of unique specificity to a particular cell type, which we explore in the next section.

The topology of the period classes recapitulated the DMD analysis of the previous section, with the left-hand cluster corresponding to one DMD mode and the right-hand cluster to another; however, the suggested time-dynamics are richer (with five periods, opposed to the two suggested by DMD) and also involve longer periods, as DMD suggested periods on the order of a few days (likely the difference between the two period classes being lumped together into a single mode), whereas each period class is 29 days or longer. The flow of information in this topology again suggests the bottleneck role of IL12-p40 in conveying cycling information to the DMD mode containing platelets from the DMD mode and containing many important growth factors, including SCF.

Integrated Data-Driven and Predictive Modeling: Th-17 Dysfunction as a Mechanism for CTP, and Simulations of Interventions

Equipped with the dimensional reduction provided by CCM, we trained statistical models for subjects A and C using their upstream nodes as predictors via LASSO regression[42] (Experimental Procedures and Supplemental Experimental Procedures). For subject A, these nodes are TPO, IL-2, IL-7, IL12-p40, ICAM1, and VCAM1. For subject C, these are TPO, nerve growth factor, C-X-C motif chemokine 1, IFN-α, IFN-β, IL-13, IL-17f, IL-18, IL-1RA, IL-22, IL-27, IL-5, IL-6, IL-9, macrophage inflammatory protein 1β (MIP1B), RANTES, FASL, and transforming growth factor-α (TGF-α).

For subject A, the error-minimizing model had predictor coefficients of comparable orders of magnitude with the exception of VCAM1. However, all the best-performing models (mean standard error of 10% with 10-fold cross-validation) gave priority to IL-2, IL12-p40, and TPO, in that order. For subject C, the best-performing models were slightly less accurate (mean squared error around 20%), and often discarded many predictors, with the same preferred features appearing in all the best-performing models: IL-17f, TGF-α, TPO, and MIP1B, in this order.

Given the significant differences between subjects A and C, it is not surprising that the best internal predictors are different. However, what is unexpected is that both sets of predictors have high specificity to the life cycle of Th-17 cells. This is especially relevant to the discussion of what causes thrombocytopenia, as recent studies suggest that Th-17 cells are critical in immune (non-cycling) thrombocytopenia.[43–45]

The dominant contribution of IL-17f in subject C makes this relationship obvious, because only Th-17 cells independently secrete IL-17. We plot comparisons of platelet counts to these leading contributors for both subjects in Figure 6, where IL-17f is shown to be strongly correlated with platelet counts in subject C. Since Th-17 cells are the only cells that create IL-17, we consider this cytokine-platelet correlation to be a direct proxy for Th-17 cell counts and/or activity levels. It is reasonable to conclude that Th-17 cells are either directly implicated in the

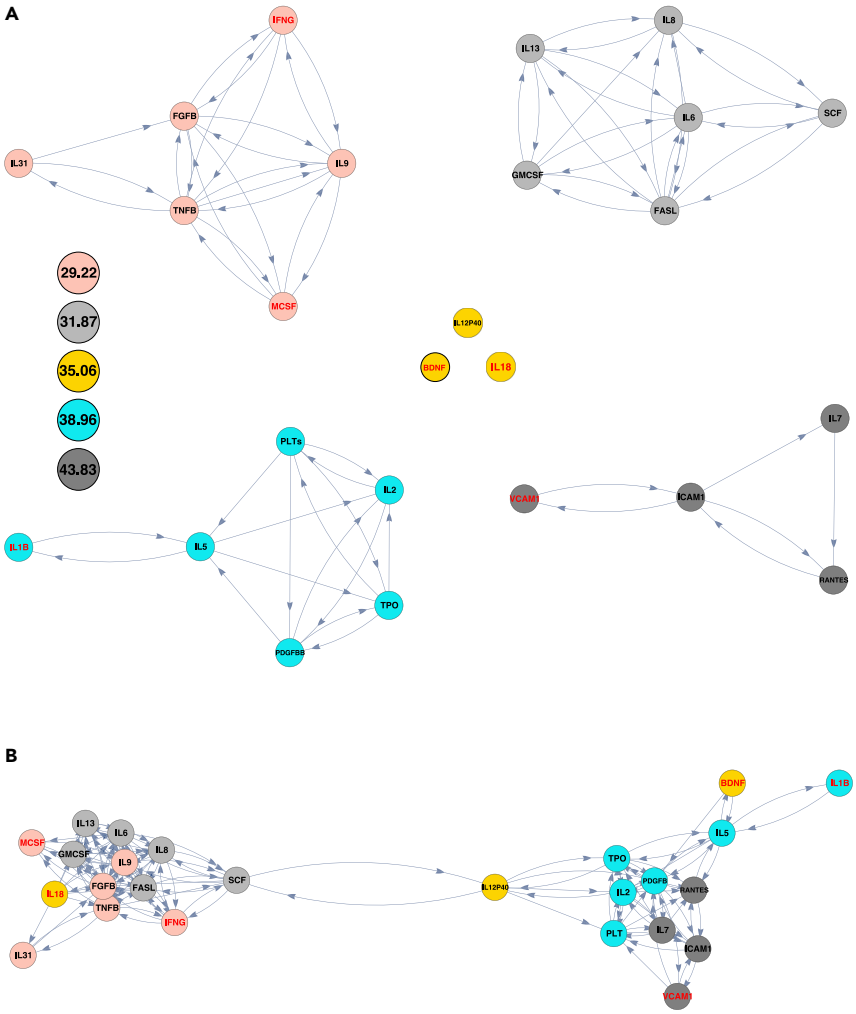


Figure 5. Mechanistic Relationships among Period Class Pairs and Cycling Cytokines in Subject A

(A) Combining the periodogram analysis and convergent cross-mapping analyses reveals a mechanistic network among elements of cytokine classes that have identical periods. Cytokines belonging to period classes are mapped in Figure S12. Cytokines are represented in the network as nodes, with node color assigned according to period (legend bottom left). Nodes with black text labels cycle with their assigned period with $\alpha = 0.05$. Nodes with red text labels cycle with their assigned period with $\alpha = 0.5$. In this panel, interactions are restricted to elements of the same periodogram class.

(B) Convergent cross-mapping analysis also uncovers links between cytokines that belong to different period classes. Here edges are colored with the same color as their source node. See Figure S13 for the equivalent in the other two subjects.

mechanism of CTP in this subject, or are directly affected by the same mechanism that is causing CTP.

In subject A, the top two predictors (IL-2 and IL-12p40) are also implicated in the life cycle of Th-17 cells, albeit in a more nuanced manner. Both cytokines are strong antagonists of polarization of Th cells to Th-17 cells.[34,35] It is interesting, then, that IL-2 correlates so strongly with platelets in both CTP-presenting subjects, despite not being selected by the models as a strong predictor for subject C. The complex role of IL-12p40 in the CTP mechanism in subject A is highlighted by the fact that IL12-p40 cycles almost completely out-of-phase with platelets (Figure 6).

As a first exploration in this direction, we performed simulations of an intervention that we believe is representative of suppressing IL-17 overexpression in subject C. To this end, we used a theoretical model of CTP[24] that was first fit to the measured data on subject C. This model was constructed directly from the mechanistic principles of thrombopoiesis, namely the differentiation of hematopoietic stem cells (HSCs) into megakaryoblasts, megakaryoblast mitosis and transformation into megakaryocytes, megakaryocyte endomitosis and shedding into platelets, and the circulating dynamics of platelets. This model has previously been shown to accurately predict CTP dynamics in multiple individuals through bifurcation and numerical analysis. We represented the pro-inflammatory IL-17, identified from the above network analysis, as leading to abnormal rates of differentiation of HSCs into the megakaryoblast lineage (κ_P in the model). This is based on the hemato-immunological observation that inflammatory and stress conditions typically lead to overproduction of HSCs and increased differentiation into the megakaryocyte and platelet lineage.[46] Using the best fit for subject C to this parameter, we then simulated the platelet and TPO levels over time for several periods in the presence of ±20% of this differentiation rate. As shown in Figure 6, this is projected to lead to an almost complete cessation of platelet fluctuations not only in subject C, but in another member of the cohort of patients plotted in Figure 1 who was not directly examined by us. This is extremely promising, as each of these two patients fits to a different κ_P as well as other kinetic parameters, yet nonetheless show the same flattening of oscillations. While the platelet volume numbers would still be considered clinically impoverished, there is a much higher rate of success with TPO mimetics in non-cycling thrombocytopenia,[47–49] and we suspect that removing the oscillatory mechanisms may be the first step in successful combination therapies.

DISCUSSION

At present, most network analyses in hematopoiesis have focused on genetic regulation at the intracellular level. However, a multitude of cytokines also transmit information as they coordinate both the production and function of blood cells. Cytokine and blood cell concentrations fluctuate daily, but additional information about the dynamic nature of the hematopoietic system

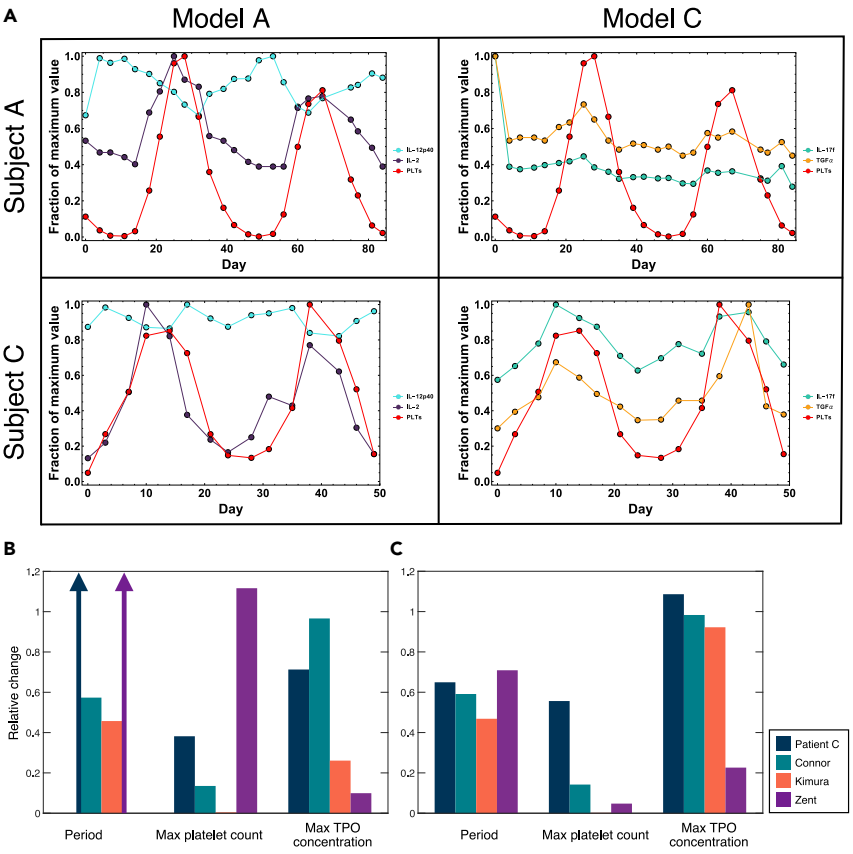


Figure 6. Leading Predictors of Platelet Levels in both CTP-Presenting Subjects

(A) Here, we show the measured platelet counts alongside cytokine counts for both subject A (top row) and subject C (bottom row). The cytokines chosen are two of the leading predictors for the LASSO fit to subject A (called model A, including IL-12p40 and IL-2, shown in the left-hand column) and the fit to subject C (called model C, including IL-17f and TGF-α, shown in the right-hand column). The model C cytokines show little dynamics in subject A, whereas one of the model A cytokines (IL-2) shows high correlation with platelets in subject C. TPO was not chosen despite being a leading predictor in both models since it is always strongly anti-correlated with platelet count. The strong correlation of IL-17f with platelets in subject/model C implicates higher Th-17 counts and/or activity when platelets are high. The anticorrelation between IL-12p40 and platelets in subject/model A suggests a more nuanced interaction between platelets and Th-17 cells, if indeed one exists. See also Figure S7.

(B) Results of *in silico* interventions for the pro-inflammatory role of IL-17 inferred in subject C, performed on subject C, as well as three other subjects for which data are available. The intervention consists of a 20% reduction in the number of stem cells entering the megakaryocyte differentiation pathway simulated using a theoretical model of CTP.[24] The y axis represents the relative change from the clinically observed values. The arrows for the Zent subject and subject C indicate that the oscillations completely vanished in the simulated trajectories.

(C) The same intervention, but with a 20% increase in the same parameter.

is revealed in oscillatory hematopoietic diseases, such as CTP. Here, we have combined information from subjects with this disease with a collection of tools uniquely positioned both to identify intervention points and untangle different phenotypes that currently are identified as only one disease—a situation which is relevant for diseases beyond CTP.

Preliminary cohort analysis of platelet and TPO concentrations from six individuals revealed that subject A had a significantly different clinical presentation than other previously studied CTP subjects and subject C in our study. Standard correlation analysis further suggested more lumping into subnetworks present in subject A than in subject C, with subject B positioned as intermediate between individuals. To explore this non-clinical phenotypic heterogeneity, we combined CCM, transfer entropy, and DMD, three non-parametric causal inference techniques from very different areas of mathematics, and found consensus on couplings uncovered in the hemato-immune networks of all three subjects. These analyses were further corroborated by searching the literature to assess the novelty of interactions revealed by our statistical and dynamical studies.

Our integrative approach uncovered many previously corroborated cytokine relationships (with 78% of the mechanistic relationships in each subject finding confirmation in the literature), with a higher percentage of supported couplings than was inferred by correlation analysis alone (57%) or by random picking of cytokine pairs and checking if there was a connection (37%). Using empirical-dynamical techniques to narrow the scope of investigation allows us to better elucidate models for each subject and to better discriminate between different phenotypes in CTP, which to our knowledge are previously unreported. For instance, it is clear from comparing subjects A and B that a particular genetic mutation (*c-mpl*) is not sufficient to induce thrombocytopenia, corroborating the knowledge that, unlike cyclic neutropenia, CTP is not easily treated by targeting a single cell type or signaling pathway.[29] From comparing both the CCM networks and periodograms for subjects A and C, it is also clear that within CTP there are different phenotypic presentations. However, these dimensionally reduced models also suggest a potential common axis rooted in control of Th cell differentiation. The seemingly atypical behavior of some regulatory cytokines in this axis (notably IL12-p40) may provide therapeutic targets.

Although these new techniques promise new understandings of this disease as well as others, some open questions are beyond our reach. Despite the father-son relationship and shared *c-mpl* mutation, the dramatic difference between subjects A and B remains puzzling, and we were not able to shed light on what induces CTP in subject A but not subject B.

We believe that the use of these tools in exploring other complex disorders is immensely promising, and has implications far

beyond CTP. Indeed, as far as the methodology is concerned, CTP represents nothing more than time series data on a system comprised of many cytokine and cell types, and so similar plots and results could be generated for any complex disorder, although we have no *a priori* evidence that it can be as successful in every case. There could be implications for our work even beyond identifying therapeutic targets and/or distinct phenotypes. For instance, the collection of corroborated relationships we uncovered, along with their strong couplings as indicated by CCM, suggests a method of more cost-effective multiplex assays by not measuring redundant cytokines in cases where they can be predicted from the levels of other cytokines with high accuracy from a pre-existing mechanistic network. Giving more rigor to this ensemble such that it could provide a robust, portable analytical tool for complex disorders will be a useful avenue of future investigation.

EXPERIMENTAL PROCEDURES

Resource Availability

Lead Contact

Morgan Craig, CHU Sainte-Justine Research Center, 3175 chemin de la Côte-Sainte-Catherine, Montréal (Québec) H3T 1C5 Canada, morgan.craig@umontreal.ca.

Materials Availability

This study did not generate new unique reagents.

Data and Code Availability

Any data which do not appear in the main text or the Supplemental Information will be made publicly available upon the publication of a companion piece in preparation, and is also available upon request. Sample code for analytical techniques is either linked to directly below or is written in pseudocode format, and can be made available in specific formats upon reasonable request.

Ethics Statement

All participants gave written informed consent. The Institutional Review Board at Stanford University approved this study (IRB no. 13735).

Cell Counts and Cytokine Assays

Blood samples were drawn from three individuals (subjects A, B, and C) with the following schedule: subject A visited the Stanford Medical Center clinic every 3–4 days over a period of 84 days and complete blood counts were taken; subject B visited every 3–10 days over a total of 83 days; and subject C visited the clinic every 3–4 days for a total of 49 days. Subject A is a 53-year-old individual with a previously described[33] presentation of CTP related to a heterozygous germ line mutation resulting in the loss of c-mpl function. Because subject B does not present with CTP and was not a patient of the clinic, his samples were collected by a separate clinic off-site and overnight-shipped to the Stanford Medical Center. Samples were received and processed on average every 3–7 days over the course of 17 samples (83 days in total). Due to the differences in sample preservation and processing time, a number (239 out of 1,054 total wells, or 22%) of the samples for subject B were flagged by quality control for having coefficient of variation (CV %) exceeding 20% between replicates. Because these could represent degraded samples, we removed these data points. No cytokine had fewer than eight valid samples remaining. To control for the missing data points and irregular sampling times, we applied cubic Hermite spline interpolation to the extant data and performed all analysis on the interpolated data. For each cytokine, we list the number of samples used and the CV% over time in the samples used (to give confidence in the accuracy of the splines, since few cytokines vary by much in this non-CTP-presenting subject) in Table S4. One sample (sample 13) in subject C was flagged and corrected using the same method. Because of the possibility that differences in sampling procedures may have adjusted overall baseline measurements between subject B and the two CTP patients, we primarily compared the internal relationships between the cytokines measured in each subject. Platelet-poor plasma was isolated to measure thrombopoietin concentration via ELISA and a panel of 62 cytokines using a Luminex immunoassay.

Lomb-Scargle Periodogram Analysis

The Lomb-Scargle periodogram[37,38] was used to estimate the period of oscillations within cell and cytokine observations for each subject. In brief, the Lomb-Scargle periodogram is a generalization of the discrete Fourier transform for unevenly spaced data that evaluates the power spectrum density of a given signal. We used the MATLAB function plomb to return to the power spectrum density and calculated the statistical alpha-level of the returned values. We noted periods detected with alpha values above 0.1 and considered those above 0.05 to be significant.

Correlations between Measures

To understand the statistical associations between platelets, thrombopoietin, and the 62 cytokines, we quantified their pairwise Pearson correlations using the cor function in R[50] and plotted the results with corrplot.[51] Hierarchical clustering based on dissimilarity using the Lance-Williams dissimilarity update formula, as implemented R,[50] was applied to group statistically similar cytokines.

Determining Causal Relationships Using CCM

CCM[11] is our primary method of use for organizing a collection of abnormalities (i.e., our 60 cytokines, platelets, and TPO), where a significant fraction are behaving unusually (oscillating) in subjects A and C, into a network of mechanistic couplings, where an edge from variable X to variable Y implies that the value of variable X has an influence on the value of the variable Y. CCM accomplishes this by transforming an important theorem in dynamical systems theory into a leave-one-out cross-validation test for mechanistic coupling. Takens' theorem[39] states that if two variables X and Y are coupled in a dynamical system, they have the property that, when embedded in the variables $(X(t), X(t-1), \ldots X(t-E))$ (a *delay embedding of order E*; same embedding for Y, so that both variables are embedded completely independently of one another), local neighborhoods are preserved—meaning if the nearest neighbors in the X embedding of observation $X(21)$ are $\{X(3), X(7), X(192)\}$, then if X and Y are coupled, the nearest neighbors of $Y(21)$ should be $\{Y(3), Y(7), Y(192)\}$. This can be rigorously tested via the following outline (full details in the Supplemental Experimental Procedures):

1. Create the order-E delay embedding of $X(t)$ and $Y(t)$(all possible E values are tested), using a portion L^* of the total length of the time series (L).
2. To test the idea of a mechanistic coupling $X \rightarrow Y$, discard a single observation $X(t^*)$(because if $X \rightarrow Y$, it is Y that contains information about X and not the converse); the discarded datum will be approximated using data from Y).
3. The $E+1$ nearest neighbors of $Y(t^*)$, $\{Y(t_1), Y(t_2), \ldots Y(t_{E+1})\}$are identified (because $E+1$ observations are necessary to triangulate a point in E-dimensional space).
4. $\{X(t_1), X(t_2), \ldots X(t_{E+1})\}$are interpolated to produce an approximation $\widehat{X}(t^*)$for the discarded observation.
5. Repeating this for every t^* produces a complete time series $\widehat{X}(t)$, which is an approximation based on assuming that $X \rightarrow Y$ holds and using Takens' embedding idea. The Pearson correlation between the real time series $X(t)$and the approximant $\widehat{X}(t)$, $\rho_{L^*}(X, \widehat{X})$, is measured.
6. The process is repeated for all $L^* \leq L$. This gives a relationship between how good the Takens' approximant is, (ρ_{L^*}), versus how much data was used, (L^*). The relevant measures from the test are (ρ_L), the strength of the test with the most data included, and also the Spearman correlation between L^* and ρ_{L^*}, because in a true coupling the slope between the amount of data we have and the amount of predictive power we have should be monotonic.

For each of the $63 \times 62 = 3{,}906$ (we have 63 total time series measured; self-maps are redundant, so that the total is not 63×63) total possible cross-maps, we calculated the cross-mapping skill as a function of library length L (which cannot exceed 24, the total number of measurements) and embedding dimension E. We then performed Spearman rank-correlation on cross-mapping skill and L for each cross-mapping to isolate those that showed CCM. An

interaction pair that provided accurate forecasting was only considered a *convergent* cross-map if the p value of the Spearman correlation was less than 0.068. This value was chosen rather than the more typical 0.05 because it is the threshold at which the interaction TPO-platelets satisfies convergence, which is a fundamental interaction in thrombopoiesis.

Surrogate/Permuted CCM Test

For the 60 interactions not previously reported in the literature, we performed additional tests. To rule out false positives due to similar periodograms, we repeated the CCM analysis with bootstrapped surrogates having identical periodograms to the original measurements, using a previously established algorithm for generating such surrogates.[40] This guarantees that each surrogate time series not only has the same primary oscillation frequency (the number labeling the period classes in Figure 5, for instance), but also contains other subdominant modes that might be present in the original data. For each interaction, we generated 99 surrogates (the original data representing the 100th candidate) and measured the CCM forecast accuracy and Spearman p value. If the original data gave a forecast accuracy in the top 5% of all convergent maps, we considered this interaction to "pass" the permutation test. We implemented the algorithm[40] via an external package, which itself was a subset of the transfer entropy package employed (see next section).

Transfer Entropy

We further cross-examined these 60 undocumented interactions using the orthogonal inference method of transfer entropy.[15] Transfer entropy bins different areas of state space into discrete, labeled chunks, and then considers a time series of a continuous variable as a signal on the discrete states. The number of discrete states is determined by the length of the time series.[52] Two time series are then considered causally related if the information, or entropy in one discrete signal is reduced given information about the other discrete signal. The difference in entropy between the signal by itself and the signal conditioned on the second variable is the transfer entropy. If the transfer entropy is significantly higher between two variables than between bootstrapped surrogates of the variables[40] (using the same bootstrapping technique described in the previous section), then we consider them to be causally related. In this work, we examined three different methods of discretizing the state space, based on methods previously useful in biological time series[41] (uniformly spaced bins, bins whose size is inferred by kernel density estimation on a normal probability distribution fit to the data, and bins whose size is learned via the Darbellay-Vajda algorithm).[53] As the length of our time series informs the number of states used,[52] four were used in our analysis. We only considered a result to be positive if it was positive using all three discretization schemes. Exact details on the algorithm as implemented and why they were chosen can be found in the Supplemental Experimental Procedures. The methods were implemented via the package that was published simultaneously with the application to biological time series.[41]

Lasso

We used MATLAB's *Lasso* (and *lassoPlot*) function (which imposes a constraint on the $\mathbf{L}_1$ norm of all fit coefficients onto a linear regression) with the direct upstream predictors of platelets in the causal networks for the three subjects (for subject B, who has no platelets in the network, we tested the model of subject A, since subject B is the father of subject A). The target to fit was the amount of platelets. For platelets and all predictors, we first employed Z score normalization such that each quantity had zero mean and unit variance. More information and error plots can be found in the Supporting Information.

DMD

DMD[16–19] is a technique for extracting eigenfunctions of the Koopman (or, "composition") operator $\mathbf{U}$, a transfer-operator-theoretic representation of a dynamical system that takes observations of variables in a system at one time point into the next, so that each column represents a "snapshot" of the whole system as observed in a single moment. This statement can be written as $\varphi(t+1) = \mathbf{U}\varphi(t) + \mathbf{R}$, where $\varphi(t)$ is the observation of all the variables in the system at time t and $\mathbf{R}$ is residual error in this linear approximation. This operator therefore offers a linear model of a nonlinear system—the operator acting on one snapshot gives the next one, so that it represents the flow of the dynamical system, and in the theoretical limit where the number of observations becomes infinite, the Koopman operator perfectly recaptures the original nonlinear system, albeit at the price of exchanging a finite, nonlinear system for an infinite-dimensional linear one. The linearity offers dimensional reduction, because the main eigenfunctions capture much of the dynamic information observed in the data. These eigenfunctions are complex exponentials, meaning they represent coupled modes of the variables that grow, shrink, and oscillate over time (in the modes revealed in our data there is no growing and shrinking, only pure oscillations). Additional information can be found in the Supplemental Experimental Procedures.

Since the operator itself is represented by the simple expression above, the technical challenge lies in picking a $\mathbf{U}$ that minimizes the residuals $\mathbf{R}$, and then extracting the eigenfunctions—each the purview of computational algebra algorithms. After Z score normalization and subtraction of the mean over the entire dataset taken over the observations, DMD was implemented via the MATLAB package provided by Marko Budisić (https://github.com/mbudisic/koopman).

Simulated Intervention along the Th-17 Axis Using the Existing Model

We simulated the proposed intervention along the Th-17 axis by periodically modulating the rate of differentiation $(\kappa_P Q^*)$ in our previously published model of (cyclic) thrombocytopenia[24] (see Figure S14 for a complete model schematic; for the governing equations, see Equations S3–S7). For each subject, we simulated a 20% reduction in differentiation lasting 1 day every 21 days for a total of 6 cycles. We then characterized the resulting change in maximum platelet/TPO concentrations and in period length, and compared these values with their original values, established through simultaneously fitting to untreated oscillating platelet and TPO concentrations, as previously described.

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.patter.2020.100138.

ACKNOWLEDGMENTS

The authors are extremely grateful to Indiana Delsart for bibliographic support. We thank Jeff Gerold, Sim Sinai, Charleston Noble, Andrei Gheorghe, Alison Hill, Joseph Dexter, Michael Mackey, and Tony Humphries for their helpful comments and suggestions. This work was supported by Discovery Grant and Discovery Launch Supplement RGPIN-2018-04546 (to M.C.), an NSERC Postdoctoral Fellowship (to M.C.), NIH grant DP5OD019851 (supporting M.C. and M.S.K.), and Human Frontiers Science Program Long-Term Fellowship LT000183/2018-L (to J.M.M.).

AUTHOR CONTRIBUTIONS

M.S.K. conceived and designed the CCM, LASSO, DMD, and transfer entropy studies as well as performed the accompanying analysis, and wrote the manuscript. J.M.M. analyzed the interaction network. H.Z. conceived and designed the sampling, cell counts, and cytokine assays, carried out the experimental work, and wrote the manuscript. M. Chien conceived and designed the sampling, cell counts, and cytokine assays, carried out the experimental work, and wrote the manuscript. J.L.Z. oversaw and wrote the manuscript. M. Craig conceived and designed the CCM study, performed the statistical analyses, and wrote the manuscript. All authors critically reviewed each version of the manuscript and approved the final version for submission.

DECLARATION OF INTERESTS

All authors declare no conflicts of interest.

Received: June 2, 2020
Revised: September 25, 2020
Accepted: October 12, 2020
Published: December 11, 2020

REFERENCES

1. Collins, A.L., Lunt, P.W., Garrett, C., and Dennis, N.R. (1993). Holoprosencephaly: a family showing dominant inheritance and variable expression. J. Med. Genet. *30*, 36–40.
2. Muenke, M., Gurrieri, F., Bay, C., Yi, D.H., Collins, A.L., Johnson, V.P., Hennekam, R.C., Schaefer, G.B., Weik, L., and Lubinsky, M.S. (1994). Linkage of a human brain malformation, familial holoprosencephaly, to chromosome 7 and evidence for genetic heterogeneity. Proc. Natl. Acad. Sci. U S A *91*, 8102–8106.
3. Varki, R., Sadowski, S., Pfendner, E., and Uitto, J. (2006). Epidermolysis bullosa. I. molecular genetics of the junctional and hemidesmosomal variants. J. Med. Genet. *43*, 641–652.
4. Janku, P., Robinow, M., Kelly, T., Bralley, R., Baynes, A., Edgerton, M.T., and Opitz, J.M. (1980). The van der Woude syndrome in a large kindred: variability, penetrance, genetic risks. Am. J. Med. Genet. *5*, 117–123.
5. Chitnis, T., Glanz, B.I., Gonzalez, C., Healy, B.C., Saraceno, T.J., Sattarnezhad, N., Diaz-Cruz, C., Polgar-Turcsanyi, M., Tummala, S., Bakshi, R., et al. (2019). Quantifying neurologic disease using biosensor measurements in-clinic and in free-living settings in multiple sclerosis. NPJ Digital Med. *2*, 123.
6. Kantarci, O.H. (2019). Phases and phenotypes of multiple sclerosis. Continuum Lifelong Learn. Neurol. *25*, 636–654.
7. Kent, J.W., Jr. (2009). Analysis of multiple phenotypes. Genet. Epidemiol. *33*, S33–S39.
8. Verma, A., Bang, L., Miller, J.E., Zhang, Y., Lee, M.T.M., Zhang, Y., Byrska-Bishop, M., Carey, D.J., Ritchie, M.D., Pendergrass, S.A., et al. (2019). Human-disease phenotype map derived from PheWAS across 38,682 individuals. Am. J. Hum. Genet. *104*, 55–64.
9. Ye, H., Beamish, R.J., Glaser, S.M., Grant, S.C.H., Hsieh, C.-h., Richards, L.J., Schnute, J.T., and Sugihara, G. (2015). Equation-free mechanistic ecosystem forecasting using empirical dynamic modeling. Proc. Natl. Acad. Sci. U S A *112* (13), E1569–E1576.
10. Chang, C.-W., Ushio, M., and Hsieh, C.-h. (2017). Empirical dynamic modeling for beginners. Ecol. Res. *32*, 785–796.
11. Sugihara, G., May, R., Ye, H., Hsieh, C.-H., Deyle, E., Fogarty, M., and Munch, S. (2012). Detecting causality in complex ecosystems. Science *338*, 496–500.
12. Ye, H., Deyle, E.R., Gilarranz, L.J., and Sugihara, G. (2015). Distinguishing time-delayed causal interactions using convergent cross mapping. Sci. Rep. *5*, 14750.
13. Clark, A.T., Ye, H., Isbell, F., Deyle, E.R., Cowles, J., Tilman, G.D., and Sugihara, G. (2015). Spatial convergent cross mapping to detect causal relationships from short time series. Ecology *96*, 1174–1181.
14. Tsonis, A.A., Deyle, E.R., Ye, H., and Sugihara, G. (2018). Convergent cross mapping: theory and an example. In Advances in Nonlinear Geosciences, A.A. Tsonis, ed. (Springer International Publishing), pp. 587–600.
15. Schreiber, T. (2000). Measuring information transfer. Phys. Rev. Lett. *85*, 461–464.
16. Schmid, P. (2010). Dynamic mode decomposition of numerical and experimental data. J. Fluid Mech. *656*, 5–28.
17. Budisic, M., Mohr, R., and Mezic, I. (2012). Applied Koopmanism. Chaos Interdisciplin. J. Nonlinear Sci. *22*, 047510.
18. Mezic, I. (2013). Analysis of fluid flows via spectral properties of the Koopman operator. Annu. Rev. Fluid Mech. *45*, 357–378.
19. Tu, J.H., Rowley, C.W., Luchtenburg, D.M., Brunton, S.L., and Kutz, J.N. (2014). On dynamic mode decomposition: theory and applications. J. Comput. Dyn. *1*, 391–421.
20. Marbach, D., Prill, R.J., Schaffter, T., Mattiussi, C., Floreano, D., and Stolovitzky, G. (2010). Revealing strengths and weaknesses of methods for gene network inference. Proc. Natl. Acad. Sci. U S A *107*, 6286–6291.
21. Marbach, D., Costello, J.C., Küffner, R., Vega, N.M., Prill, R.J., Camacho, D.M., Allison, K.R., Aderhold, A., Bonneau, R., Chen, Y., et al. (2012). Wisdom of crowds for robust gene network inference. Nat. Methods *9*, 796–804.
22. Cohen, T., and Cooney, D.P. (1974). Cyclic thrombocytopenia. Case report and review of literature. Scand. J. Haematol. *12*, 9–17.
23. Kimura, F., Nakamura, Y., Sato, K., Wakimoto, N., Kato, T., Tahara, T., Yamada, M., Nagata, N., and Motoyoshi, K. (1996). Cyclic change of cytokines in a patient with cyclic thrombocytopenia. Br. J. Haematol. *94*, 171–174.
24. Langlois, G.P., Craig, M., Humphries, A.R., Mackey, M.C., Mahaffy, J.M., Belair, J., Moulin, T., Sinclair, S.R., and Wang, L. (2017). Normal and pathological dynamics of platelets in humans. J. Math. Biol. *75*, 1411–1462.
25. von Schulthess, G.K., and Gessner, U. (1986). Oscillating platelet counts in healthy individuals: experimental investigation and quantitative evaluation of thrombocytopoietic feedback control. Scand. J. Haematol. *36*, 473–479.
26. Apostu, R., and Mackey, M.C. (2008). Understanding cyclical thrombocytopenia: a mathematical modeling approach. J. Theor. Biol. *251*, 297–316.
27. Belair, J., and Mackey, M.C. (1987). A model for the regulation of mammalian platelet production. Ann. N.Y Acad. Sci. *504*, 280–282.
28. Santillán, M., Mahaffy, J.M., Belair, J., and Mackey, M.C. (2000). Regulation of platelet production: the normal response to perturbation and cyclical platelet disease. J. Theor. Biol. *206*, 585–603.
29. Langlois, G.P., Arnold, D.M., Potts, J., Leber, B., Dale, D.C., and Mackey, M.C. (2018). Cyclic thrombocytopenia with statistically significant neutrophil oscillations. Clin. Case Rep. *6*, 1347–1352.
30. Dale, D., and Mackey, M. (2015). Understanding, treating and avoiding hematological disease: better medicine through mathematics? Bull. Math. Biol. *77*, 739–757.
31. Connor, D.E., and Joseph, J.E. (2011). Cyclic thrombocytopenia associated with marked rebound thrombocytosis and fluctuating levels of endogenous thrombopoietin and reticulated platelets: a case report. Am. J. Hematol. 120–122.
32. Zent, C.S., Ratajczak, J., Ratajczak, M.Z., Anastasi, J., Hoffman, P.C., and Gewirtz, A.M. (1999). Relationship between megakaryocyte mass and serum thrombopoietin levels as revealed by a case of cyclic amegakaryocytic thrombocytopenic purpura. Br. J. Haematol. *105*, 452–458.
33. Zhang, H., Hou, Y., Brar, R.S., Zhang, B., Chen, Z., Abidi, P., Jin, J., Gotlib, J.R., and Zehnder, J.L. (2016). Identification of a novel MPL loss of function mutation in a patient with cyclic thrombocytopenia and characterization of this syndrome. Blood *128*, 1376.
34. Adler, E.M. (2007). IL-2 antagonizes Th17 differentiation. Sci. Signal. *2007*, tw103.
35. Lim, H.X., Hong, H.-J., Jung, M.Y., Cho, D., and Kim, T.S. (2013). Principal role of IL-12p40 in the decreased Th1 and Th17 responses driven by dendritic cells of mice lacking IL-12 and IL-18. Cytokine *63*, 179–186.
36. Bruin, M., Tijssen, M.R., Bierings, M., and de Haas, M. (2005). Juvenile cyclic amegakaryocytic thrombocytopenia: a novel entity. J. Pediatr. Hematol. Oncol. *27*, 148–152.
37. Scargle, J.D. (1982). Studies in astronomical time series analysis. II. Statistical aspects of spectral analysis of unevenly spaced data. Astrophys. J. *263*, 835–853.
38. Lomb, N.R. (1976). Least-squares frequency analysis of unequally spaced data. Astrophys. Space Sci. *39*, 447–462.
39. Takens, F. (1981). "Detecting Strange Attractors in Turbulence," Lecture Notes in Mathematics, *898* (Berlin Springer Verlag), p. 366.
40. Schreiber, T., and Schmitz, A. (2000). Surrogate time series. Phys. Nonlinear Phenomena *142*, 346–382.
41. Lee, J., Nemati, S., Silva, I., Edwards, B.A., Butler, J.P., and Malhotra, A. (2012). Transfer entropy estimation and directional coupling change detection in biomedical time series. BioMedical Eng. *11*, 19.
42. Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. J. R. Stat. Soc. Ser. B (Methodological) *58*, 267–288.

43. Ji, L., Zhan, Y., Hua, F., Li, F., Zou, S., Wang, W., Song, D., Min, Z., Chen, H., and Cheng, Y. (2012). The ratio of Treg/Th17 cells correlates with the disease activity of primary immune thrombocytopenia. PLoS One *7*, 1–9.

44. Ye, X., Zhang, L., Wang, H., Chen, Y., Zhang, W., Zhu, R., Fang, C., Deng, A., and Qian, B. (2015). The role of IL-23/Th17 pathway in patients with primary immune thrombocytopenia. PLoS One *10*, 1–13.

45. Zhang, G., Zhang, P., Liu, H., Liu, X., Xie, S., Wang, X., Wu, Y., Chang, J., and Ma, L. (2017). Assessment of Th17/Treg cells and Th cytokines in an improved immune thrombocytopenia mouse model. Hematology *22*, 493–500.

46. Pietras, E.M. (2017). Inflammation: a key regulator of hematopoietic stem cell fate in health and disease. Blood *130*, 1693–1698.

47. Bussel, J.B., Cheng, G., Saleh, M.N., Psaila, B., Kovaleva, L., Meddeb, B., Kloczko, J., Hassani, H., Mayer, B., Stone, N.L., et al. (2007). Eltrombopag for the treatment of chronic idiopathic thrombocytopenic purpura. N. Engl. J. Med. *357*, 2237–2247.

48. Kuter, D.J. (2009). Thrombopoietin and thrombopoietin mimetics in the treatment of thrombocytopenia. Annu. Rev. Med. *60*, 193–206.

49. Zeng, Y., Duan, X., Xu, J., and Ni, X. (2011). TPO receptor agonist for chronic idiopathic thrombocytopenic purpura. Cochrane Database Syst. Rev. *7*, https://doi.org/10.1002/14651858.CD008235.pub2.

50. R Core Team, R. (2017). A Language and Environment for Statistical Computing (R Foundation for Statistical Computing).

51. Wei, T., and Simko, V. (2017). R Package "corrplot": Visualization of a Correlation Matrix, Version 0.84.

52. Ramos, A., and Macau, E. (2017). Minimum sample size for reliable causal inference using transfer entropy. Entropy *19*, 150.

53. Darbellay, G.A., and Vajda, I. (1999). Estimation of the information by an adaptive partitioning of the observation space. IEEE Trans. Inf. Theor. *45*, 1315–1321.