Interpreting the deluge of omics data: new approaches offer new possibilities

Aarash Bordbar

Sinopia Biosciences, San Diego, CA, United States of America

Dear Sir,

Developments in assay technologies have transformed biology from a data-poor qualitative field to a data-rich quantitative field. The cost of biological data generation has exponentially decreased in the past two decades, allowing simultaneous measurement of several types of biomolecules, including DNA, RNA, proteins, and metabolites. Particularly for Transfusion Medicine, studies have generated metabolomics and proteomics data of stored red blood cells (RBCs) and platelets detailing time-dependent changes. These high-throughput, or "omics", datasets measure hundreds of metabolites and proteins at tens of time points for each replicate, resulting in studies with over 10,000 data points. As seen in other fields that have generated omics datasets for nearly two decades, it is apparent that data analysis and interpretation, rather than data generation, is becoming the bottleneck of biological discovery[1].

The need for sophisticated data analysis is evident from many of the recently published metabolomics studies on the RBCs and platelet storage legion. Such studies are encyclopedic in nature, often only validating existing biochemical findings previously found using low throughput assays. With so many significant changes, it can be daunting to ipinpoint which metabolite(s) are indicative of a diagnostic biomarker or are important to change for better preservation of red cells. For example, in a recently published metabolomics dataset for RBCs stored in SAGM additive solution, 75% of the over 100 measured metabolites significantly changed over the storage period[2]. How can such high-dimensional data be conceptualised and used for biological discovery and intervention?

Systems biology and bioinformatics are disciplines for deciphering the complex interactions of cells at a global level using high-throughput datasets. In these fields, many statistical and mechanistic methods exist for analysing datasets and modelling cellular processes, each with their own advantages and disadvantages. Researchers should choose techniques based on the cellular process in question and the number of experimental datasets available. Using these considerations, we developed a systems biology platform that: 1) determined RBCs undergo three metabolic states during storage, 2) identified metabolic biomarkers that differentiate between the metabolic states, and 3) pinpointed key regions of cellular metabolism to experimentally perturb that will likely alter and prolong the earlier metabolic states (Figure 1).

First, we measured 100+ intracellular and extracellular metabolites over time for RBCs stored in SAGM additive solution using absolute quantitative metabolomics[2]. Using statistical analyses, we reduced the high dimensional data set. In particular, principal component analysis was used to determine that the cells undergo three distinct metabolic phases during storage. To our surprise, more traditional *in vitro* haematologic parameters currently used for quality assessment, such as haemolysis, ion and gas concentrations, and RBC indices, were less indicative of elapsed storage time. Thus, the biochemical signature we identified indicative of the storage lesion may act as a better quantitative *in vitro* measure of RBC quality, allowing for better benchmarking of novel RBC processing techniques.

Second, the principal component analysis-derived biochemical signature for the storage lesion was further refined by selecting a smaller but "core" subset of the 100+ metabolites. In subsequent work, we showed that tracking just eight metabolites could accurately assess which of the three metabolic states the RBCs are in[3]. Further, the signature was found to be general as the work was experimentally validated in different labs using different mass spectrometry techniques, different RBC processing methods, and different additive solutions (SAGM and AS3).

Third, we have developed comprehensive metabolic networks of RBCs and platelets to assess the changes in metabolic pathway usage during storage. Metabolic network reconstructions are an aggregation of the available primary literature from low- and high-throughput datasets representing the current knowledge of a particular cell's metabolism. These networks are suited for modelling metabolism at the cellular scale[4] and even for analysing blood metabolomics datasets from individuals[5]. For example, the RBC network was constructed from 60+ research articles and textbooks published over the past 50 years. The network contains 292 metabolic reactions catalysed by 281 enzymes. Metabolic networks are analogous to flow networks, in which metabolites "flow" through the network similar to liquids flowing in a pipe. By integrating the time-course measurement of a large portion of metabolites during storage with the underlying "pipeline" of the RBC metabolic network, the status of flow throughout the network can be computed for each of the three metabolic states. Similar to the large number of significant changes seen for the metabolites, we found that the activity of roughly 50% of metabolic reactions significantly change as RBCs shift from one metabolic state to the next. In order to determine important and potentially causal changes to the RBC storage legion, we developed and applied network modelling algorithms that determine

Blood Transfus 2017; **15**: *189-90* DOI 10.2450/2017.0356-16

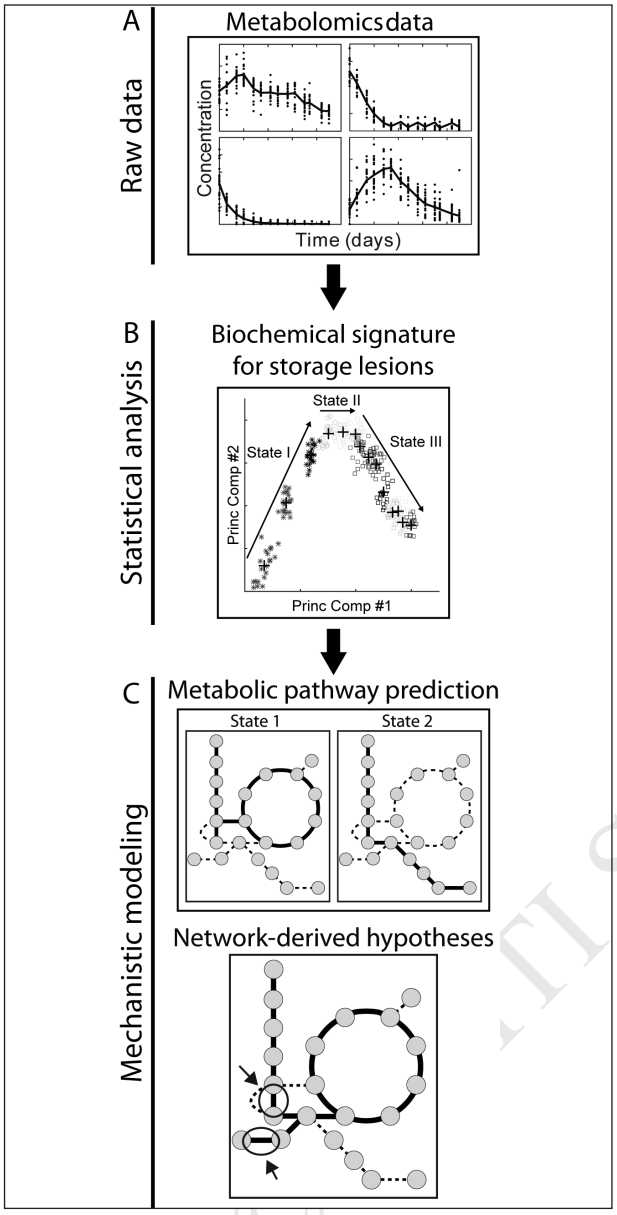

Figure 1 - Workflow of a systems biology platform for analysing high-throughput data in Transfusion Medicine.
(A) Absolute quantitative metabolomics data are generated for stored red blood cells and platelets. (B) Statistical analyses, such as principal component analysis, are used to define a biochemical signature for the storage lesions and for determining the time intervals of each metabolic state during storage. The signature can be used for *in vitro* assessment of blood cell age and a benchmark for new processing methods. (C) Metabolomics data can be integrated with metabolic networks to determine pathway usage for each metabolic state. Network modeling algorithms can then be applied to generate hypotheses on experimental perturbations to help minimise pathway usage changes that occur in subsequent metabolic states. A toy example is shown in which through two perturbations the second metabolic state is reverted to a state similar to the first.

a minimal set of metabolic reactions that if modified would globally alter and prolong the first and second metabolic states.

Using the computationally derived metabolic "lynchpins" of the RBC storage lesion, we have devised perturbation strategies to assess our ability to alter and prolong the first and second metabolic states, thus avoiding the later state of the storage lesion. In collaboration with Drs. D'Alessandro and Hansen at the University of Colorado in Denver, these computationally derived hypotheses will be tested. Metabolomics data will be generated to assess if the interventions affect the metabolic state, a quantitative *in vitro* benchmark for improving red cell quality. Successful interventions will be tested for positive effects on non-metabolic markers of the RBC storage lesion, such as micro-particle generation and cellular rheology. Further, predictions found to be inaccurate allow the reconciliation of the generated high-throughput data with the network structure for discovering new biology not previously captured in the literature.

Systems biology and bioinformatics methodologies provide statistical and mechanistic methods to interpret high-dimensional datasets. A metabolic network's structure represents the compilation of the currently available knowledge of a particular cell's metabolism. Integrating datasets from stored RBCs and platelets will provide a novel opportunity to interpret and harness these unwieldy datasets.

Acknowledgements

These projects are supported by the National Heart, Lung, and Blood Institute (R43HL127843 and R44HL123074) and Europe Research Council (232816).

Disclosure of conflicts of interest

AB is a co-founder of Sinopia Biosciences.

References

1) D'Alessandro A, Kriebardis AG, Rinalducci S, et al. An update on red blood cell storage lesions, as gleaned through biochemistry and omics technologies. Transfusion 2015; **55**: 205-19.
2) Bordbar A, Johansson PI, Paglia G, et al. Identified metabolic signature for assessing red blood cell unit quality is associated with endothelial damage markers and clinical outcomes. Transfusion 2016; **56**: 852-62.
3) Paglia G, D'Alessandro A, Rolfsson Ó, et al. Biomarkers defining the metabolic age of red blood cells during cold storage. Blood 2016; **128**: e43-50.
4) Bordbar A, Monk JM, King ZA, et al. Constraint-based models predict metabolic and associated cellular functions. Nat Rev Genet 2014; **15**: 107-20.
5) Bordbar A, McCloskey D, Zielinski DC, et al. Personalized whole-cell kinetic models of metabolism for discovery in genomics and pharmacodynamics. Cell Systems 2015; **1**: 283-92.

Arrived: 12 December 2016 - Revision accepted: 15 December 2016
Correspondence: Aarash Bordbar
Sinopia Biosciences
600 W Broadway Suite 700
San Diego, CA 92101, USA
e-mail: abordbar@sinopiabio.com