



HAL
open science

A causal discovery approach for streamline ion channels selection to improve drug-induced TdP risk assessment

Safaa Al-Ali, Jordi Llopis-Lorente, Maria Teresa Mora, Maxime Sermesant, Beatriz Trénor, Irene Balelli

► To cite this version:

Safaa Al-Ali, Jordi Llopis-Lorente, Maria Teresa Mora, Maxime Sermesant, Beatriz Trénor, et al.. A causal discovery approach for streamline ion channels selection to improve drug-induced TdP risk assessment. 2023 Computing in Cardiology (CinC), Oct 2023, Atlanta (GA), United States. 10.22489/CinC.2023.009 . hal-04105144

HAL Id: hal-04105144

<https://hal.science/hal-04105144>

Submitted on 24 May 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A causal discovery approach for streamline ion channels selection to improve drug-induced TdP risk assessment

Safaa Al-Ali¹, Jordi Llopis-Lorente², Maria Teresa Mora², Maxime Sermesant¹, Beatriz Trénor², Irene Balelli¹

¹ Centre INRIA d'Université Côte d'Azur, EPIONE team, Sophia Antipolis, France

² Centro de Investigación e Innovación en Bioingeniería (Ci2B), Universitat Politècnica de València, Valencia, Spain

Abstract

The query of causality is of paramount importance in biomedical data analysis: assessing the causal relationships between the observed variables allows to improve our understanding of the tackled medical condition and better support decision-making. Torsade de Pointes (TdP) is an extremely serious drug-induced cardiac side effect, which can provoke ventricular fibrillation and lead to sudden death. TdP is related to abnormal repolarizations in single cells, and the minimum set of ion channels needed to correctly assess TdP risk is still an open question. Discovering causal relations between drug-induced ionic channels' perturbations could shed new light on the underlying mechanisms leading to TdP, and drive variable selection to improve TdP-risk assessment. In this work, we propose to apply the causal discovery method ICA-Linear Non-Gaussian Acyclic Model (ICA-LiNGAM) to uncover the relations across the 7 ion channels identified by the Comprehensive in vitro Proarrhythmia Assay (CiPA) initiative as potentially related to the induction of TdP: I_{Kr} , I_{Na} , I_{NaL} , I_{CaL} , I_{K1} , I_{Ks} and I_{to} . The obtained Bayesian causal network can be explored to infer the downstream impact of the ionic currents in the drug's safety label. We consider 109 drugs of known torsadogenic risk (51 unsafe) listed by CredibleMeds. We identify I_{Kr} , I_{NaL} and I_{CaL} as the ones which directly affect TdP-risk assessment, and suggest that I_{Na} perturbations could potentially have a high impact on proarrhythmic risk induction. Our causality-based results were further confirmed by independently performing binary drug risk classification, which shows that the combination of the 3 selected ions maximizes the classification accuracy and specificity, outperforming state-of-the-art approaches based on alternative ion channel combinations.

1. Introduction

Identifying causal links between variables of importance takes a leading position in facilitating stable inference and making rational decisions in several areas, by overcoming the bias induced through standard correlation-based statistical and machine-learning approaches. For instance, in medicine, understanding the causal relationship between risk factors and disease can help develop effective prevention and treatment strategies and prioritize the available information to improve risk assessment tools. Causal discovery is a branch of causal research, whose aim is to learn the cause-effect relationships of observational variables, typically through causal graphs. Several methods have been proposed to learn such graphs, for instance by exploring several assumptions on the independent causal function representing the relation between each variable (or node) and its parents' nodes in the graph, or the independent variable-specific noise term [1]. Among them, a classic approach is provided by the linear non-Gaussian acyclic model (LiNGAM) [2], which assumes that the causal relations between variables are simply linear and their noise term is non-Gaussian to enforce the identification of the causal directions. An extension of the LiNGAM algorithm, ICA-LiNGAM, further assumes that the observed variables are not independent due to the presence of unobserved or latent variables [3]. In ICA-LiNGAM, independent component analysis (ICA, [4]) is used to separate the observed variables into their independent components, hence LiNGAM is applied.

Drug-induced torsade de pointes (TdP) is one of the most frightening drugs' side effects since it can trigger ventricular fibrillation and ultimately lead to sudden death. TdP is closely associated with abnormal repolarization, hence by a prolongation of the QT interval. The human ether-à-go-go-related gene (hERG) is responsible for the rapid component of the delayed rectifier current (I_{Kr}) which is one of the major repolarizing currents in the heart [5]. Proarrhythmic risk assessment of drugs is tradition-

ally based on the evaluation of the hERG channel and the measure of the delayed ventricular repolarization on the electrocardiogram (ECG), i.e. the QT interval prolongation [6]. This method can accurately classify high TdP-risk drugs through a single analytical assessment considering a unique ionic channel and exclusively focusing on ventricular repolarization [7]. However, it has been observed (eg [8]) that I_{Kr} blocking may be not sufficient to assess drug-induced TdP risk and is prone to produce false positives. Consequently, recent studies have been proposed to incorporate additional ion channels other than I_{Kr} for a more reliable drug safety assessment [9–11]. Several ion channels’ combinations have been retained in the literature as the most relevant to assess drug-induced TdP risk such as I_{Kr} , I_{CaL} , I_{Na} [9, 12] or I_{Kr} , I_{CaL} , I_{NaL} , and I_{Ks} [10].

In this work, we propose to use the ICA-LiNGAM causal discovery approach for upstream ion channels’ selection, to improve TdP risk assessment. The rest of the paper is organized as follows: in Section 2, we introduce the causal discovery algorithm employed in this work and the considered drug data-set. In Section 3, we present the results obtained for uncovering the relationships between the ion channels and the binary drugs’ label as well as their straightforward applicability for the classification task. Section 4 provides our conclusions and discussion of future directions for the presented work.

2. Materials and method

2.1. Drugs data-set

A total of 109 drugs from CredibleMeds [13] with known torsadogenic risk (37 with known risk, 14 with possible risk, 13 with conditional risk and 45 with no proven risk) were used in this study. More details about the data set are available in [10] and the supplementary material herein. In this work, we consider a binary classification task labeling drugs of confirmed or possible TdP-risk as unsafe, and drugs with conditionally or no proven TdP risk as safe.

For every drug, we consider two pharmacological data: the IC_{50} for each of the seven ionic currents addressed by the CiPA initiative, and the effective free therapeutic plasma concentration (EFTPC), defined as the drug concentration in the plasma required to produce the desired therapeutic effect in the body. As input of the causal algorithm, we use the ion channels blocked fraction, here denoted by Bf_{Ion} :

$$Bf_{Ion} := \left[1 + \left(\frac{IC_{50}Ion}{EFTPC} \right)^h \right]^{-1}, \quad (1)$$

where h denotes the Hill coefficient, the number of drug molecules assumed to be sufficient to block a ion channel.

2.2. ICA-LiNGAM causal discovery algorithm

ICA-LiNGAM [3] is a function-based causal discovery algorithm [14], which extends the Linear Non-Gaussian Additive Model (LiNGAM) [2]. LiNGAM assumes that the causal relations between observed variables can be represented by a non-Gaussian linear acyclic model, i.e. a directed acyclic graph (DAG) [15] where the causal functions between each variable and its parents’ nodes are linear, and the error term follows a non-Gaussian distribution. Let us denote by $\mathbf{X} = \{x_i, 1 \leq i \leq n\}$ the set of observed variables: we can encode the DAG structure through its adjacency matrix $\mathbf{B} = \{b_{ij}\}$, where b_{ij} represents the strength of the connection between the variables x_i and x_j . The matrix \mathbf{B} could be permuted by simultaneous permutations of equal rows and columns to be strictly lower triangular, in accordance with the acyclicity assumption. Finally, denoting by Pa_i the set of parents nodes of x_i (i.e. the nodes who causally precede x_i), the generating process for variable $x_i \in \mathbf{X}$ writes:

$$x_i = \sum_{x_j \in Pa_i} b_{ij}x_j + e_i, \quad (2)$$

where $e_i \in \mathbf{E}$ denotes a centered non-Gaussian noise with non-zero variance. In a matrix form, we can rewrite Equation (2) as follows:

$$\mathbf{X} = (\mathbb{I} - \mathbf{B})^{-1}\mathbf{E}, \quad (3)$$

\mathbb{I} being the identity matrix. In this work, we consider the ICA-LiNGAM method, an extension of LiNGAM based on the assumption that observed variables may be dependent due to the presence of unobserved or latent variables [3]. Indeed, Equation (3) defines the independent component analysis model since the noise terms e_i are independent and non-Gaussian. ICA [4] is used estimate the mixing matrix $\mathbf{A} = (\mathbb{I} - \mathbf{B})^{-1}$. Further, a set of permutations and scaling of the obtained independent parameters are performed to estimate the causal order k and the adjacency matrix as well. The ICA-LiNGAM steps are summarized in Algorithm 1.

3. Results

In Figure 1 we show the causal graph obtained by applying ICA-LiNGAM to the ion channels blocked fractions, for the seven considered ionic currents that have been proven to be potential for TdP induction: I_{Kr} , I_{Na} , I_{NaL} , I_{CaL} , I_{K1} , I_{Ks} and I_{to} . We introduce an extra *Label* node to investigate the relations of the blockade parameters with the known proarrhythmic risk. We used 5-fold cross-validation over the 109 drugs, and the final graph represented in Figure 1 accounts for the occurrences of causal

Algorithm 1 ICA-LiNGAM

Input: Data matrix \mathbf{X} **Output:** $\tilde{\mathbf{B}}$, estimation of \mathbf{B} (Equation (3))

1. Given \mathbf{X} , use ICA to obtain the decomposition $\mathbf{X} = \mathbf{A}\mathbf{S}$, where \mathbf{S} contains the independent components. Define $\mathbf{W} := \mathbf{A}^{-1}$.
 2. Compute the matrix $\tilde{\mathbf{W}}$, the unique permutation of \mathbf{W} without any leading diagonal zero.
 3. Compute $\tilde{\mathbf{W}}'$, which corresponds to $\tilde{\mathbf{W}}$ after dividing each row by its corresponding diagonal element.
 4. Compute an estimation $\hat{\mathbf{B}}$ of $\mathbf{B} := \mathbf{I} - \tilde{\mathbf{W}}'$.
 5. Estimate the causal order or the parents variables for each variable $x_i \in \mathbf{X}$ by finding the permutation matrix $\tilde{\mathbf{P}}$ of $\hat{\mathbf{B}}$ such that $\tilde{\mathbf{B}} := \tilde{\mathbf{P}}\hat{\mathbf{B}}\tilde{\mathbf{P}}^T$ is as close as possible to a strictly lower triangular structure. This can be measured using $\sum_{i \leq j} \tilde{\mathbf{B}}_{ij}^2$.
 6. Return the permuted matrix $\tilde{\mathbf{B}}$.
-

arrows identified in each fold. One can see that the Kr channel is systemically identified as a direct cause leading to the drug label, which is directly caused as well by CaL and NaL ionic currents. Despite the fact that channel Na is not affecting the node label directly, it still plays an important role since it is causally related to Kr , CaL , and NaL channels.

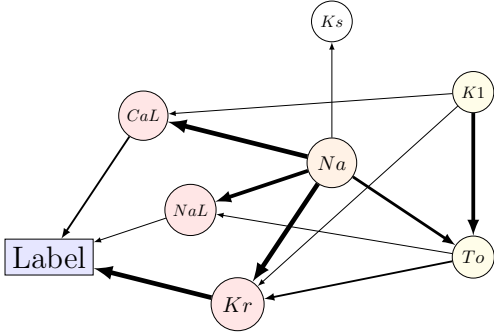


Figure 1: Causal graph generated by ICA-LiNGAM, performing 5-fold cross-validation. The arrows' thickness represents the number of occurrences of the causal arrows over the 5 splits.

Next, we perform an independent binary classification of the 109 drugs used for this study. In particular, we used two classical machine learning classifiers: RandomForest and K-Nearest-Neighbors, whose prescriptions are finally combined through a majority voting classifier [16]. A 5-fold cross-validation was performed as well for the classification task.

In Figure 2 we present the accuracy and speci-

ficity scores for the ion channels combinations revealed by the obtained causal graph (Figure 1), starting from the most stably identified parent of the label node Kr , up to its farthest ancestors. The highest accuracy levels are achieved by the two combinations (Kr, CaL, NaL) and (Kr, CaL, NaL, Na) . For the ion-combination (Kr, CaL, NaL) we obtained a mean accuracy level achieving 93.59%, whereas the combination (Kr, CaL, NaL, Na) allows to reach a mean accuracy of 92.68%.

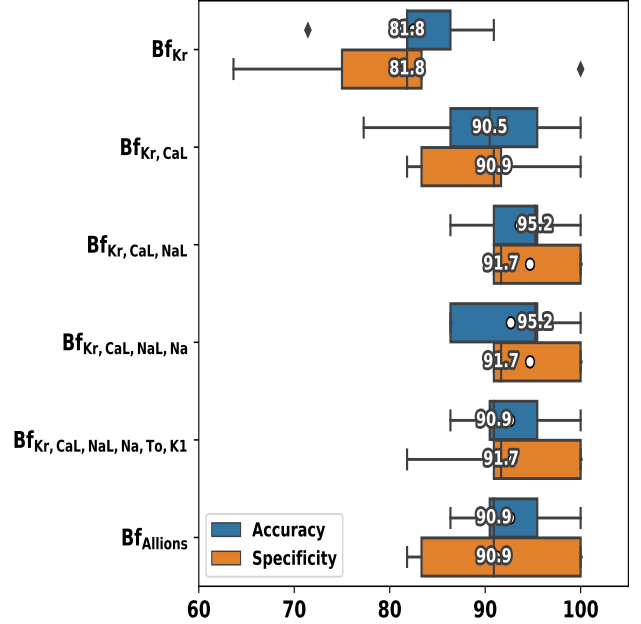


Figure 2: Accuracy and specificity levels of drug's safety classification based on the combinations of ions as suggested by the ICA-LiNGAM algorithm (Figure 1) using the voting classifier with RandomForest and K-Nearest-Neighbors. The white dots correspond to the mean values and the numerical values plotted over each box represent the median.

Furthermore, our causality-based classification results show a high mean specificity value reaching 94.70% for (Kr, CaL, NaL) . There is no further improvement beyond these combinations, indicating that the causal graph has been effective in identifying the main parameters which directly inform on the drug safety classification, namely (Kr, CaL, NaL) .

Table 1 summarizes the drugs safety classification performance of our causally-based method with respect to some state-of-the-art methods [9, 12]. Our ion-combination selection shows the best mean AUC and specificity scores, i.e. 0.94 and 94.7% respectively, outperforming the other methods. These results emphasize the importance of tak-

ing into consideration ion channels *CaL* and *NaL* in addition to *Kr* to improve drug induced TdP-risk classification.

Ion combination	AUC	Acc.	Sens.	Spec.	Ref.
<i>Kr, CaL, Na</i>	0.91	90.9%	88%	87%	[9]
<i>Kr, CaL, Na</i>	NA	87%	73%	89%	[12]
<i>Kr, CaL, NaL</i>	0.94	93.59%	92%	94.7%	Our method

Table 1: Mean values of the performance (AUC, accuracy, sensitivity and specificity) of our causality-based classification procedure compared to some state-of-the-art methods which use ion channels for proarrhythmic risk classification.

4. Conclusion

The study presented here highlights the importance of using causal discovery methods to infer ion channels selection for the drug-induced TdP risk. The obtained results suggest that I_{Kr} , I_{CaL} and I_{NaL} are the most critical ions to be considered for TdP risk assessment. These results are consistent with previous works (e.g., [10]) where the authors showed the crucial role of these ions in the computation of *in silico* arrhythmogenic biomarkers proposed for TdP risk assessment. In addition, the proposed causal approach can bring valuable insights concerning the downstream effects of perturbation of other ionic channels, and the obtained causal graph can further be deployed to infer the safety of new compounds. The current study has the potential to expand its scope by incorporating additional biological variables such as action potential [17], or *in-silico* biomarkers proposed in the literature [10]. This would significantly aid in selecting parameters prior to simulations, resulting in reduced computational time and a more reliable and rationale variable selection.

Acknowledgements

This work was supported by the European Union’s Horizon 2020 Research and Innovation Program SimCardioTest project which has received funding from the European Union’s Horizon 2020 research and innovation program under grant agreement No. 101016496.

References

- [1] Chen W, Cai R, Zhang K, Hao Z. Causal discovery in linear non-gaussian acyclic model with multiple latent confounders. *IEEE Transactions on Neural Networks and Learning Systems* 2021;33(7):2816–2827.
- [2] Shimizu S, Hoyer PO, Hyvärinen A, Kerminen A, Jordan M. A linear non-gaussian acyclic model for causal discovery. *Journal of Machine Learning Research* 2006;7(10).
- [3] Shimizu S, Inazumi T, Sogawa Y, Hyvarinen A, Kawahara Y, Washio T, Hoyer PO, Bollen K, Hoyer P. Directlingam: A direct method for learning a linear non-gaussian structural equation model. *Journal of Machine Learning Research JMLR* 2011;12(Apr):1225–1248.
- [4] Stone JV. Independent component analysis: an introduction. *Trends in cognitive sciences* 2002;6(2):59–64.
- [5] Nerbonne JM, Kass RS. Molecular physiology of cardiac repolarization. *Physiological reviews* 2005;85(4):1205–1253.
- [6] Witche HJ. Drug-induced hERG block and long QT syndrome. *Cardiovascular therapeutics* 2011;29(4):251–259.
- [7] Guideline IHT. The non-clinical evaluation of the potential for delayed ventricular repolarization (QT interval prolongation) by human pharmaceuticals. S7B http://www.ich.org/products/guidelines/safety/articlesafety_guidelines.html 2005;.
- [8] Li Z, Ridder BJ, Han X, Wu WW, Sheng J, Tran PN, Wu M, Randolph A, Johnstone RH, Mirams GR, et al. Assessment of an *in silico* mechanistic model for proarrhythmia risk prediction under the CiPA initiative. *Clinical Pharmacology Therapeutics* 2019;105(2):466–475.
- [9] Kramer J, Obejero-Paz CA, Myatt G, Kuryshv YA, Bruening-Wright A, Verducci JS, Brown AM. Mice models: superior to the hERG model in predicting torsade de pointes. *Scientific reports* 2013;3(1):1–7.
- [10] Llopis-Lorente J, Gomis-Tena J, Cano J, Romero L, Saiz J, Trenor B. *In silico* classifiers for the assessment of drug proarrhythmicity. *Journal of Chemical Information and Modeling* 2020;60(10):5172–5187.
- [11] Mistry HB. Complex versus simple models: ion-channel cardiac toxicity prediction. *PeerJ* 2018;6:e4352.
- [12] Zhou X, Qu Y, Passini E, Bueno-Orovio A, Liu Y, Vargas HM, Rodriguez B. Blinded *in silico* drug trial reveals the minimum set of ion channels for torsades de pointes risk assessment. *Frontiers in Pharmacology* 2020;1643.
- [13] Wosley R, Romero K, Heise W. Risk categories for drugs that prolong QT and induce torsades de pointes (TdP), 2019.
- [14] Nogueira AR, Pugnana A, Ruggieri S, Pedreschi D, Gama J. Methods and tools for causal discovery and causal inference. *Wiley interdisciplinary reviews data mining and knowledge discovery* 2022;12(2):e1449.
- [15] Foraita R, Spallek J, Zeeb H. *Directed Acyclic Graphs*. New York, NY: Springer New York, 2014; 1481–1517.
- [16] Ruta D, Gabrys B. Classifier selection for majority voting. *Information Fusion* 2005;6(1):63–81. ISSN 1566-2535. Diversity in Multiple Classifier Systems.
- [17] Jeong DU, Yoo Y, Marcellinus A, Lim KM. Application of convolutional neural networks using action potential shape for *in-silico* proarrhythmic risk assessment. *Biomedicine* 2023;11(2):406.

Address for correspondence:

Safaa Al-Ali
 Centre INRIA d’Université Côte d’Azur, EPIONE team, 06902
 Sophia Antipolis, France.
safaa.al-ali@inria.fr