A causal discovery approach to streamline ionic currents selection to improve drug-induced TdP risk assessment

Safaa Al-Ali¹, Jordi Llopis-Lorente², Maria Teresa Mora², Maxime Sermesant¹, Beatriz Trenor², 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 ion 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 Pro-arrhythmia 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 on the drug’s safety label. We consider 109 drugs of known torsadogenic risk listed by CredibleMeds. We identify $I_{Kr}$, $I_{NaL}$ and $I_{CaL}$ as the ones that directly affect TdP-risk assessment, and suggest that $I_{Na}$ perturbations could potentially have a high impact on pro-arrhythmic 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 ionic currents 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 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, the linear non-Gaussian acyclic model (LiNGAM) [2] 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 latent variables [3]: 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 drug 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 traditionally 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 ion channel and exclusively focusing on ventricular repolarization \cite{7}. However, it has been observed (e.g., \cite{8}) that $I_{Ks}$ 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 multiple ion channels, since drug-altered currents may interact and cause the disrupted cardiac electrophysiology, for a more reliable drug safety assessment \cite{9–11}. Several ion channels’ combinations have been retained in the literature as the most relevant to assess drug-induced TdP risk, such as $I_{Kr}$, $I_{CaL}$, $I_{NaL}$ \cite{9,12} or $I_{Kr}$, $I_{CaL}$, $I_{Nat}$, and $I_{Ks}$ \cite{10}.

In this work, we propose to use ICA-LiNGAM for upstream ion channels’ selection, to improve TdP risk assessment. We introduce the causal discovery algorithm ICA-LiNGAM and the considered drug data set. In Section 3, we present our results: the uncovered relationships between the ion channels and the binary drugs’ label, and 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 \cite{13} with known torsadogenic risk (37 with known, 14 with possible, 13 with conditional, and 45 with no proven risk) were used in this study. More details about the data set are available in \cite{10} and the supplementary material herein. We consider a binary classification task, labeling drugs of confirmed or possible TdP-risk as unsafe, and drugs with conditionally or no proven risk as safe.

For every drug, we use 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 concentration ($EFTPC$), defined as the drug concentration ($C_i$) that is needed to achieve the therapeutic effect.

2.2. ICA-LiNGAM algorithm

ICA-LiNGAM \cite{3}, summarized in Algorithm 1, is a function-based causal discovery algorithm \cite{14}, which extends the Linear Non-Gaussian Additive Model (LiNGAM) \cite{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) \cite{15} where the causal functions between each variable and its parents’ nodes are linear, and the error term follows a centered non-Gaussian distribution with non-zero variance. Let $X = \{x_i, i = 1, \ldots, n\}$ be the set of observed variables: we can encode the DAG structure through its adjacency matrix $B = \{b_{ij}\}$, where $b_{ij}$ represents the strength of the connection between the variables $x_i$ and $x_j$. The matrix $B$ could be permuted 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 $x_i \in X$ writes:

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

where $\epsilon_i \in E$ are the error terms. We can rewrite Equation (2) in a matrix form,

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

$I$ being the identity matrix. The ICA-LiNGAM method considered here is based on the additional assumption that observed variables may be dependent due to the presence of unobserved latent variables \cite{3}. Indeed, Equation (3) defines the independent component analysis model since observed variables may be dependent due to the presence of unobserved latent variables \cite{3}. ICA \cite{4} is used to estimate the mixing matrix $A = (I - B)^{-1}$. Further, a set of permutations and scaling of the obtained independent parameters are performed to obtain the causal order $k$ and the adjacency matrix as well.

**Algorithm 1 ICA-LiNGAM**

**Input:** Data matrix $X$

**Output:** $B$, estimation of $B$ (Equation (3))

1. Given $X$, use ICA to obtain the decomposition $X = AS$, where $S$ contains the independent components. Define $W := A^{-1}$.
2. Compute the matrix $\widehat{W}$, the unique permutation of $W$ without any leading diagonal zero.
3. Compute $\widehat{W}'$, which corresponds to $\widehat{W}$ after dividing each row by its corresponding diagonal element.
4. Compute an estimation $\widehat{B}$ of $B := I - \widehat{W}'$. $\widehat{B}$ is as close as possible to a strictly lower triangular structure ($\sum_{i \leq j} \widehat{B}_{ij}^2$).
5. Estimate the causal order for each variable $x_i \in X$ by finding the permutation matrix $P$ of $\widehat{B}$ such that $\overline{BPB}$ is as close as possible to a strictly lower triangular structure ($\sum_{i \leq j} \overline{B}_{ij}^2$).
6. Return the permuted matrix $\overline{B}$.

3. Results

In Figure 1 we show the causal graph obtained by applying ICA-LiNGAM to the ion channels blocked fractions,
for the seven ionic currents identified by the CiPA initia-
tive: $I_{Kr}$, $I_{Na}$, $I_{CaL}$, $I_{K1}$, $I_{Ks}$ and $I_{to}$. We introduce an
extra Label node to investigate the relations of the block-
ade parameters with the known pro-arrhythmic risk. We
used 5-fold cross-validation over the 109 drugs, and the
final graph represented in Figure 1 accounts for the occur-
rences of causal arrows identified in each fold. A directed
causal arrow from variable $A$ to variable $B$ indicates that
changes in the parent node $A$ can directly cause changes
in the children node $B$. One can see that the $I_{Kr}$ channel
is systemically identified as a direct cause leading to the
node label, which is directly caused as well by $I_{CaL}$ and $I_{NaL}$
channels. Despite the fact that $I_{Na}$ channel is not affecting
the node label directly, it still plays an important role since
it is causally related to $I_{Kr}$, $I_{CaL}$, and $I_{NaL}$ channels. Indeed,
the obtained DAG suggests that if a drug has an impact on
$I_{Na}$ blockade, it will also have an impact on its descendant
ion channels blockade, $I_{Kr}$, $I_{CaL}$, $I_{NaL}$.

Next, we perform an independent binary classification
of the 109 drugs used for this study. In particular, we
used two classical machine learning classifiers: Random-
Forest 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 specificity 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, $I_{Kr}$,
up to its farthest ancestors. The highest accuracy levels are
achieved by the two combinations ($I_{Kr}$, $I_{CaL}$, $I_{NaL}$) and ($I_{Kr}$,
$I_{CaL}$, $I_{NaL}$, $I_{Na}$). For the ion-combination ($I_{Kr}$, $I_{CaL}$, $I_{NaL}$)
we reach a mean accuracy level of 93.59%, whereas the combination ($I_{Kr}$, $I_{CaL}$, $I_{NaL}$, $I_{Na}$) allows to achieve a mean
accuracy of 92.68%.

![Figure 1: Causal graph generated by ICA-LiNGAM, through 5-fold cross-validation. The arrow thickness represents the number of occurrences of the causal arrow over the 5 splits.](image)

Figure 2: Accuracy and specificity of drugs’ safety classification based on the combinations of ionic currents suggested by the DAG in Figure 1, using the voting classifier with Random-Forest 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 shows a
high mean specificity, reaching 94.70% for ($I_{Kr}$, $I_{CaL}$, $I_{NaL}$).
There is no further improvement beyond this combination,
indicating that the causal graph has been effective in iden-
tifying the main parameters which directly inform the drug
safety classification, namely ($I_{Kr}$, $I_{CaL}$, $I_{NaL}$). These results
are consistent with the importance of the main outward
current $I_{Kr}$ to assure cardiac repolarization and the counter-
acting effects of inward currents $I_{CaL}$ and $I_{NaL}$ to maintain
repolarization reserve and prevent proarrhythmic events.

Table 1 summarizes the drug safety classification perfor-
mance 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. Unlike cardiac simulation classifiers, our ap-
proach does not require any cardiac action potential model
to assess drug TdP risk: it only depends on the pore block
model based on IC$_{50}$ and Hill coefficients for various ion
channels to calculate drug blockade. These results empha-
size the importance of taking into consideration ion chan-
nels $I_{CaL}$ and $I_{NaL}$ in addition to $I_{Kr}$ to improve drug-induced
TdP-risk classification.

For completeness’ sake, we have tested the efficacy of
our method for a 4-class classification (see Sec. 2.1), which reveals that \((I_{Kr}, I_{CaL})\) is preferable for this task, along with a great heterogeneity across classes: this deserves further analysis, and is left as future work.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(K_r, CaL, Na)</td>
<td>0.91</td>
<td>90.3%</td>
<td>82%</td>
<td>87%</td>
<td>[9]</td>
</tr>
<tr>
<td>(K_r, CaL, NaL)</td>
<td>NA</td>
<td>87%</td>
<td>73%</td>
<td>89%</td>
<td>[12]</td>
</tr>
<tr>
<td>(K_r, CaL, NaL)</td>
<td>0.94</td>
<td>93.59%</td>
<td>92%</td>
<td>94.7%</td>
<td>Our method</td>
</tr>
</tbody>
</table>

Table 1: Mean values of the performance (AUC, accuracy, sensitivity, and specificity) of our causality-based classification approach compared to some state-of-the-art methods which use ion currents for pro-arrhythmic 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 \(I_{Na}\) are the most critical ion currents 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 ionic currents 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 currents, 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, for a reduced computational time and a more reliable and rational 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


Address for correspondence:

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