

Creating a Kidney Transplant Risk Calculator using GEO datasets (Group 24)

460352996 470066919 480145820 480407614

GitHub code repository is [here](#)

This version was compiled on May 31, 2020

Kidney failure is the final stage of renal disease and is dangerous for the body as the excretory system fails to function properly. Therefore medical intervention in the form of renal dialysis or organ transplantation is required. Organ transplantation is a lifesaving treatment for those people diagnosed with Kidney disease and is greatly preferred over renal dialysis. It has the potential to offer a better quality of life for the patient - fewer health problems and reduced restrictions on diet- to name a few benefits. Despite the benefits, organ allocation has posed itself as a major resource balance problem. This shortage in donor kidneys results in the need for a careful assessment in allocating organs to patients that potentiate in maximal survival time. We propose a tool that can aid in the effective and accurate allocation of these organs.

Transplant | Regression | Prediction | Immunosuppression

Introduction.

Objective. Our main objective is to create a risk calculator that can assist practitioners in their decision making for patient kidney allocation and inform prescriptions for immunosuppressive drugs.

Target Audience. We believe that our product is best suited for clinical settings that foster shared decision making between the patient and the nephrologist. According to Gordan's 2013 paper, Shared Decision Making promotes patient centered care allowing for the integration of the nephrologist's expertise on renal allograft dysfunction as well as the patient's values and beliefs concerning future treatment. Our tool will provide an opportunity for discussion between the patient and practitioner that concerns the nature of treatment pre, during and post organ transplantation.

Risk Calculator.

1. Predicting acute rejection. Firstly, our calculator will utilise patient's genetic expression from RNA-Seq to predict the chance of them experiencing T-cell/antibody mediated acute rejection, which is a sudden decline in graft function. This can help practitioners in deciding whether more aggressive immune suppression methods are required to prevent rejection.

1.1. Data cleaning and processing Acute Rejection (AR) calculator is based on data from GSE120396, GSE120649, GSE131179 from Gene Expression Omnibus. We combined the three datasets to build part one of the risk calculator. For two of the them, we performed cpm, log2 transformation appropriately and changed ensemble id to gene symbols. We joined the three datasets based on common gene symbols and performed quantile normalisation to reduce batch effect. This allowed more robust predictions that can be generalised to different sequencing platforms. We performed feature selection based on differentially expressed genes between stable and acute rejection patients, using the limma function. The top 100 most significant genes were selected for a more comprehensive model.

1.2. Selecting a Model For model selection, penalised logistic regression models were selected as they allow output of a specific

risk rather than a binary outcome, and it reduces issues associated with overfitting, thus solving the large p small n problem. Between ridge, lasso and elastic net, elastic net was selected as it shows the highest accuracy, a stable AUC, and small Brier score using cross validation.

1.3. Output In our product, the input raw counts will be transformed. If the data uses ENSEMBL ID, we converted them to official gene symbols. The output will be a colour coded gauge plot showing both percentage of the patient's risk and percentage of the population risk of acute rejection. This visualisation is clear and easy to understand which is suitable for our target audience.

1.4. Limitations If the input doesn't include the same genes in the model, a new model will be trained on the spot. However this creates a limitation of the long computational time.

2. Estimating Time to de novo DSA Presence. Survival curve (from 'Donor Mismatches' tab) is based on data provided by Dr. Germain Wong. Here we use the data provided by dr Germaine wong to calculate an estimate predicting the outcome of a graft based on the recipient's phenotypic information.

We will then estimate the probability until donor specific antibodies arise for the patient phenotype, based upon potential numbers of eplet mismatches they could have with donor kidneys. DSA presence has been found to be associated with many forms of rejection and graft failure.

Eplets are small arrangements of amino acid residues that are polymorphic. That is they occur in several different forms on Human Leukocyte antigen (HLA) molecules. Eplets are essential components of HLA epitopes (part of an antigen molecule where an antibody attaches itself) that are recognised by the antibodies

HLA (Human Leukocyte antigen) are cell surface proteins, responsible for regulating the immune system in humans. They are important in graft rejection: where any cell displaying some other HLA type is seen as an invader by the body's immune system = rejection of the tissue that bears those cells.

2.1. Data cleaning and processing Based on a study by Dayoub et al. in 2018 we choose age and gender as phenotypic predictors for graft survival. Here we first create a survival object using presence or absence of class 2 de nova DSAs and the number of days for it to form, which is our dependent variable for our Kaplan Meier curve. here we focused on DSA as its an established biomarker for predicting antibody-mediated rejection which in turn is the leading cause of graft loss

2.2. Output The output is a plot estimating the survival probability of a graft based on the patient's age and gender. As the patient would not know the proportion of mismatches he/she would have with a particular donor we stratified the mismatches based on the mean number of mismatches in the provided data which was 30. This gives the practitioner and the patient a better idea of what to expect based on what the previous patient with similar characteristics experienced. Ideally, this can be used in resource allocation and prescribing medication. Just to highlight, this is not used as a predictor of graft survival but only a guide.

2.3. Limitations The main issue was the size of the data set is small. (only 198 entries). So when we stratify it further based on age, gender, and mismatches we end up with little information for each subgroup. therefore we don't have enough data to represent the entire subpopulation fairly, meaning that some curves are uninformative.

3. Predicting Operational Tolerance. A way to minimise graft dysfunction is for the patient to take immunosuppressive drugs, which as a consequence could increase their risk of contracting diseases. However, some rare patients are found to be operationally tolerant, which means that they still maintain stable graft function after removal of these drugs. Hence, our final part will be assessing the patient's reliance on immune suppression, which can help with prescription making for the practitioner.

3.1. Data cleaning and processing To predict operational tolerance, we collected a dataset that contained raw gene expressions from tolerant patients, and normal patients which still required immune suppression for graft function. After quantile normalisation and batch effect removal, we performed feature selection. Firstly, genes with noticeable counts in at least one groups was retained by using the edgeR package. We then selected the most differentially expressed genes between the two groups using multiple t-tests. Finally, a review by Massart et al. (2017) suggested a collection of genes that were highly differential between tolerant and normal patients, and so these were also added to our final training dataset.

3.2. Selecting a Model Similar to part 1, we performed penalised logistic regression models to account for possible overfitting in a large p small n situation. Elastic net once again proved to be a better model than Ridge and LASSO, showing higher accuracies and AUC after repeated cross-validation.

3.3. Output The output is a simple colour coded gauge plot showing both percentage of the patient's reliance and percentage of the population requiring immunosuppression, making it suitable for our target audience.

3.4. Limitations However, since operational tolerance is rare, only a handful of data samples are available, making it difficult to diagnose. Also, like Part 1, another limitation of the model is it has to be 'retrained' for if the input does not contain the same genes in the model.

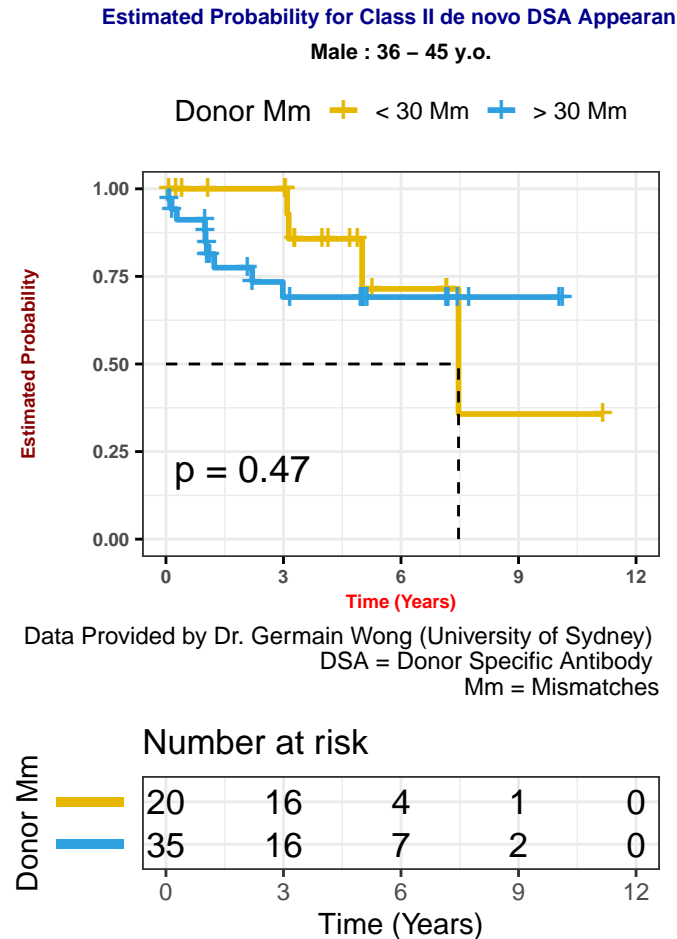
Conclusion. In conclusion, using the Shared decision-making approach for our risk calculator, we hope both the practitioner and patient are able to make informed decisions and derive the optimal treatment option. This includes allocating reliable donor kidneys to the patient and prescribing the correct dosage of immunosuppressive drugs. In terms of future work, we need more diverse and larger datasets to account for variation and possible confounders within the population, and also increase confidence. For example, the eplet data is made up of mostly Caucasian patients only. Ultimately, shared decision making should be implemented more widely within organ transplantation.

Reference List.

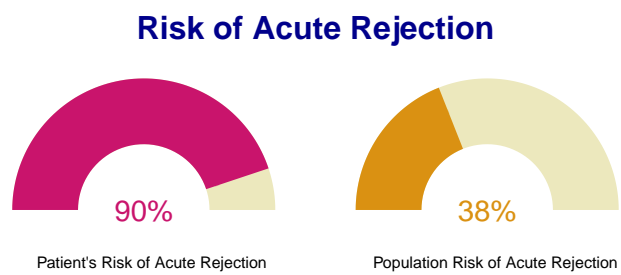
- 1.. Dayoub, J.C. Cortese, F, Anzic, A, 2018, The Effects of Donor Age on Organ transplants: a review and implications for aging research, Experimental Gerontology, Vol 110, pp. 230-240, Retrieved from <>
- 2.. Dorr, C. R., Oetting, W. S., Jacobson, P. A., & Israni, A. K. (2018). Genetics of acute rejection after kidney transplantation. Transplant International, 31(3), 263-277.
- 3.. Edgar R., Domrachev M., Lash AE. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res, 30(1),207-10.
- 4.. Gordon, E.J., 2013, Opportunities for Shared Decision Making in Kidney Transplantation, American Journal of Transplantation, Vol.13, no.5 pp. 1149-1158.
- 5.. Kleinbaum, D.G., 2005 Klein, M., 'Introduction to Survival Analysis' in Survival Analysis: A self learning text, Springer, New York, NY pp.1-43.
- 6.. Massart, A., Ghisdal, L., Abramowicz, M., & Abramowicz, D. (2017). Operational tolerance in kidney transplantation and associated biomarkers. Clinical & Experimental Immunology, 189(2), 138-157.
- 7.. Tambur, A.R. 2018, HLA-epitope Matching or Eplet Risk Stratification: The Devil is in the Details, Front Immunol, Vol.9

Appendixes.

Appendix A. Kaplan-Meier curve: Estimated Probability for Class II de novo DSA Appearance



Appendix B. Penalised Logistic Regression: Risk of Acute Rejection



Appendix C. Penalised Logistic Regression: Reliance on Immunosuppression

