Inferred Homologous Recombination Pathway Activity Predicts Survival in Breast Cancer Patients

Kenneth M.K. Mark[1,2] , Yue Wang[1,3], and Chao Cheng[1,2]

[1] Department of Genetics, Geisel School of Medicine at Dartmouth College, [2] Norris Cotton Cancer Center, Dartmouth-Hitchcock Medical Center,
[3] Huazhong University of Science and Technology, Wuhan, Hubei, China.

Introduction

Mutations in *BRCA1* or *BRCA2* have been widely known to significantly increase the risk of developing breast cancer[1]. BRCAness describes the phenomena of a subset women who also have a significantly increased risk of developing breast cancer, but don't harbor a mutated *BRCA1/2*[1]. These women are thought to have mutations or defects within their Homologous Recombination (HR) pathway, which *BRCA1/2* heavily participate[1]. HR is the primary pathway in the repair of double-stranded DNA breaks. The inefficiency or inability of cells to repair DNA through the high-fidelity method of HR, can lead to significant DNA damage and genomic instability perpetuating cancer development[1]. Recently, Peng et al performed single gene knockdown (KD) experiments of different HR pathway genes to determine the extent of resulting genomic changes[2]. Leveraging these KD profiles we have developed a technique to infer HR pathway activity based on similarity of a given breast cancer patient's gene expression profile to one of the provided KD profiles. We aim to determine if HR pathway activity (HRS) is indicative of breast cancer patient survival, specific breast cancer subtypes, and correlates with an accumulation of somatic mutations (indicating an inablity to resolve DNA damage).

Statistical Methods

In this project, I utilize both Weibull distribution and Cox's proportional hazards models to determine the association between survival time and inferred HR pathway activity (HRS) from the `survreg` and `coxph` functions within the **survival** library in R.Multiple testing correction via the Benjamani-Holm method was applied when calculating the different profiles using the `p.adjust` function. I also adjusted for *BRCA1/2* mutational status in the regression models which could confound results. HRS high and low groups were split at the mean, and their survival was graphically plotted using a Kaplan-Meier curve, with differences between their survival times assessed from a log-rank test. These were calculated by using the `survfit` and `survdiff` functions respectively from the R **survival** package. I performed linear regression, utilizing categorical dummy variables, to determine an association between HRS and intrinsic breast cancer subtypes as well as binomial logistic regression to determine an association between various hormone receptor statuses and HRS. Finally a wilcoxon-rank sum was used to determine if there was a signficant difference in the number of cumulative somatic mutations between HRS high and low groups using the `wilcoxon.test` function within R.

Results

Is there an association between HRS and survival?

Peng *et al*[2] provided gene-expression profiles of different HR pathway genes including *BRCA1*, *RAD51*, *PTEN*, and double KD of *PTEN* and *BRCA1*. We inferred HR pathway activity for Curtis[3] breast cancer patients using the BASE[4] algorithm, by determining similarity to each of the KD profiles.

	survreg (p-val)	survreg (adj)	coxph (p-val)	survreg (adj)
BRCA1(B)	*4.72 e-19*	*2.36 e-18*	*<2.2 e-16*	*<2.2 e-16*
RAD51	*9.80 e-17*	*2.45 e-16*	*<2.2 e-16*	*<2.2 e-16*
PTEN	*1.04 e-03*	*2.73 e-03*	*5.21 e-03*	*8.68 e-03*
BRCA1(A)	*5.94 e-03*	*7.43 e-03*	*2.38 e-02*	*2.98 e-02*
PTEN+BRCA1	*5.76 e-01*	*5.76 e-01*	*4.33 e-01*	*4.33 e-01*

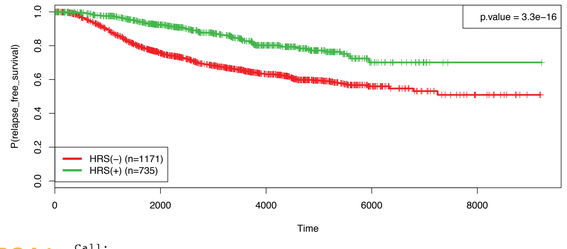
Does this remain signficant after adjusting for *BRCA1* or *BRCA2* mutational status?

RAD51

```
Call:
coxph(formula = Surv(t.rfs, e.rfs) ~ RAD51 + BRCA1.mut + BRCA2.mut,
    data = xx)

  n= 1906, number of events= 491
   (3 observations deleted due to missingness)

              coef exp(coef)  se(coef)      z Pr(>|z|)
RAD51     -0.027210  0.973157  0.003069 -8.865   <2e-16 ***
BRCA1.mut  0.013942  1.014040  0.305406  0.046    0.964
BRCA2.mut  0.324690  1.383602  0.281280  1.154    0.248
---
Signif. codes:  0  ***  0.001  **  0.01  *  0.05  .  0.1     1
```

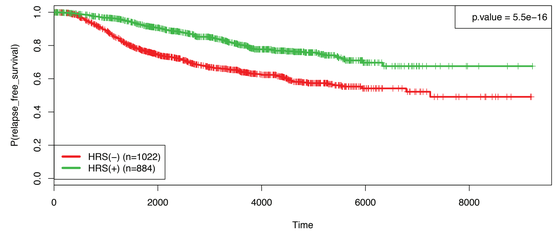


BRCA1

```
Call:
coxph(formula = Surv(t.rfs, e.rfs) ~ BRCA1 + BRCA1.mut + BRCA2.mut,
    data = xx)

  n= 1906, number of events= 491
   (3 observations deleted due to missingness)

              coef exp(coef)  se(coef)      z Pr(>|z|)
BRCA1     -0.045593  0.955430  0.004966 -9.181   <2e-16 ***
BRCA1.mut  0.050395  1.051686  0.305410  0.165    0.869
BRCA2.mut  0.374308  1.453985  0.281205  1.331    0.183
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```



Is there an association between HRS and intrinsic subtype?

```
Call:
lm(formula = xx$RAD51 ~ factor(pamr) + BRCA1.mut + BRCA2.mut)

Residuals:
    Min      1Q  Median      3Q     Max
-36.754  -7.807  -2.517   7.906  48.600

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)           10.772      6.108   1.764  0.07798 .
factor(pamr)Basal    -15.702      6.164  -2.547  0.01094 *
factor(pamr)LumA      12.482      6.133   2.035  0.04196 *
factor(pamr)LumB      -8.422      6.142  -1.371  0.17043
factor(pamr)Her2     -14.325      6.184  -2.316  0.02064 *
factor(pamr)Normal    16.503      6.204   2.660  0.00788 **
BRCA1.mut             -2.188      2.479  -0.883  0.37750
BRCA2.mut              4.239      2.594   1.634  0.10234
```

Is there an association between HRS and hormone receptor histology?

ER

```
Call:
glm(formula = ER.stat ~ xx$RAD51 + BRCA1.mut + BRCA2.mut, family = "binomial")

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.3758   0.0695   0.3333   0.9087   1.2257

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) 0.906282   0.061505  14.735   <2e-16 ***
xx$RAD51    0.087636   0.006301  13.907   <2e-16 ***
BRCA1.mut   0.032675   0.421197   0.078    0.938
BRCA2.mut   0.031107   0.471079   0.066    0.947
```

PR

```
Call:
glm(formula = PR.stat ~ xx$RAD51 + BRCA1.mut + BRCA2.mut, family = "binomial")

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.2188  -0.9987   0.5622   1.1088   1.4821

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.249535   0.054333  -4.593 4.38e-06 ***
xx$RAD51     0.038118   0.002801  13.608  < 2e-16 ***
BRCA1.mut    0.346252   0.356014   0.973    0.331
BRCA2.mut    0.207642   0.368380   0.564    0.573
```

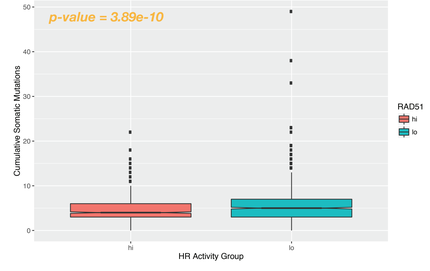
HER2

```
Call:
glm(formula = PR.stat ~ xx$RAD51 + BRCA1.mut + BRCA2.mut, family = "binomial")

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.2188  -0.9987   0.5622   1.1088   1.4821

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.249535   0.054333  -4.593 4.38e-06 ***
xx$RAD51     0.038118   0.002801  13.608  < 2e-16 ***
BRCA1.mut    0.346252   0.356014   0.973    0.331
BRCA2.mut    0.207642   0.368380   0.564    0.573
```

Do low HRS patients accumulate more somatic mutations?



Conclusions

We determined that there is indeed a significant association between inferred HR pathway activity and survival time especially for the genes *BRCA1* and *RAD51*. We determined that this association remains significant after adjusting for *BRCA1/2* mutational status and that higher HRS scores are indicative of longer relapse-free survival times. We also determined that HRS is associated with several intrinsic subsets, and that higher HRS indicates Luminal A and Normal-like samples, whereas low HRS indicated Basal-like and Her2 patients (who historically do worse). Additionally, HRS is strongly associated with each hormone receptor histology and higher HRS scores indicate positive receptor statuses (which generally yield more treatable cancers). Finally, we determined that low HRS patients, those that are unlikely to be able to correct DNA damage, have a significantly higher distribution of cumulative somatic mutations, indicating that our inferred HR pathway activity metric is sensitive to true HR activity.

Acknowledgements

I would like to thank Y.W. for providing data and helpful comments during this process.

** All code and analysis for this project is available online at: https://github.com/KmkMark/QBS121_Project

This work was supported by the American Cancer Society Research Grant IRG-82-003-30, the NIH Centers of Biomedical Research Excellence (COBRE) grant GM103534, and the Dartmouth Clinical and Translational Science Institute, under award number UL1TR001086 and KL2TR001088 from the National Center for Advancing Translational Sciences, and Dartmouth College Norris Cotton Cancer Center Support Grant P30CA023108.

References

[1] Turner, N., Tutt, A. & Ashworth, A. Hallmarks of 'BRCAness' in sporadic cancers. Nat Rev Cancer 4, 814-819, doi:10.1038/nrc1457 (2004).

[2] Peng, G. et al. Genome-wide transcriptome profiling of homologous recombination DNA repair. Nat Commun 5, 3361, doi:10.1038/ncomms4361 (2014).

[3] Curtis, C. et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature 486, 346-352, doi:10.1038/nature10983 (2012).

[4] Cheng, C., Yan, X., Sun, F. & Li, L. M. Inferring activity changes of transcription factors by binding association with sorted expression profiles. BMC Bioinformatics 8, 452, doi:10.1186/1471-2105-8-452 (2007).

[5] Pereira, B. et al. The somatic mutation profiles of 2,433 breast cancers refines their genomic and transcriptomic landscapes. Nat Commun 7, 11479, doi:10.1038/ncomms11479 (2016).