Prognostic relevance of CD73 in high-grade, late-stage ovarian carcinoma

Gregory M. Chen

June 12, 2015

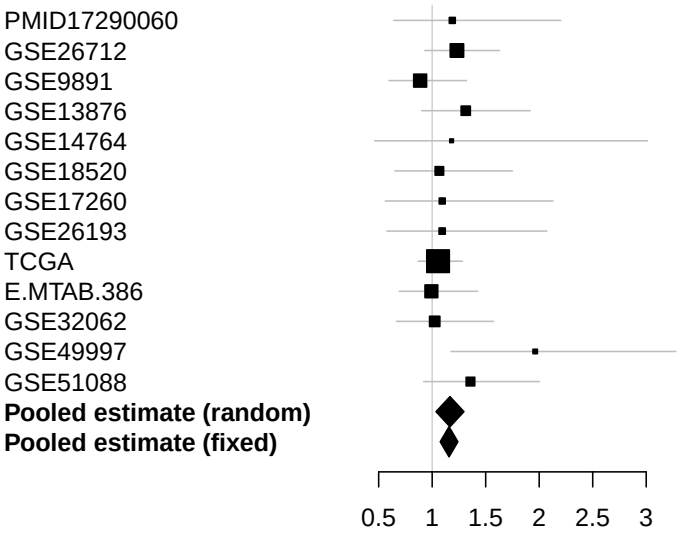
1 Datasets

1581 patients with late-stage, high-grade serous ovarian cancer were pooled from MetaGxOvarian, representing 13 datasets:

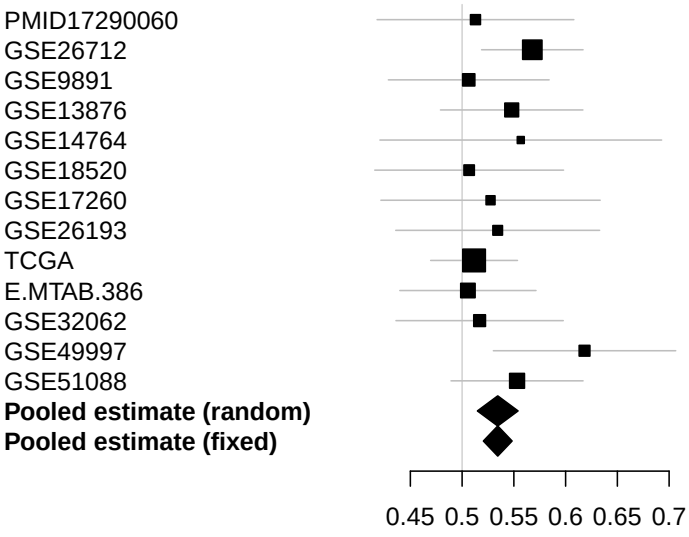
Dataset name	Number of samples
PMID17290060	59
GSE26712	185
GSE9891	140
GSE13876	98
GSE14764	41
GSE18520	53
GSE17260	43
GSE26193	47
TCGA	452
E.MTAB.386	128
GSE32062	129
GSE49997	122
GSE51088	84

2 Meta-analysis

The d- and c-indices in the forest plots below were generated by setting risk as positively associated with CD73 expression level - *i.e.*, higher expression, higher risk.



Concordance indices:



3 Survival curves

All 1581 samples in the pooled dataset had expression levels for CD73. CD73 expression levels were separated into tertiles.

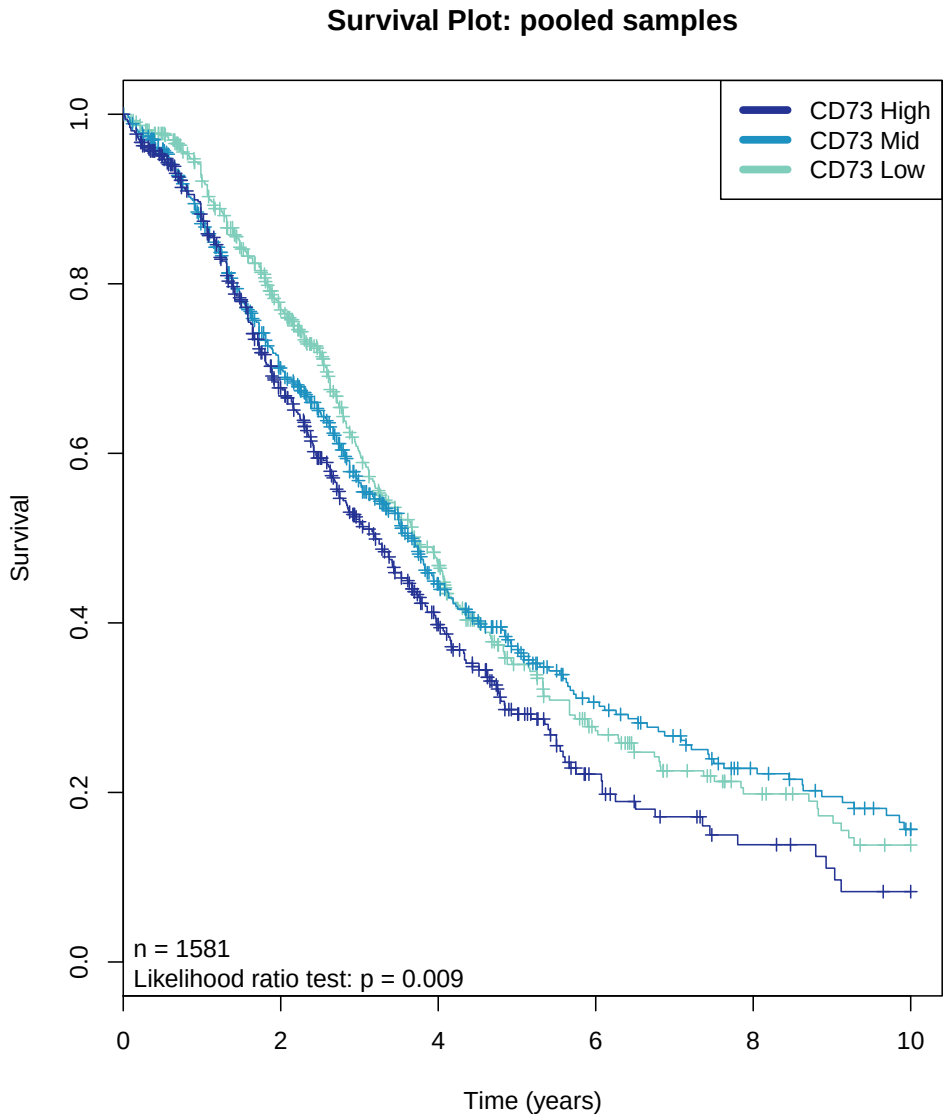


Figure 1: Survival curves for CD73 tertiles, all patients

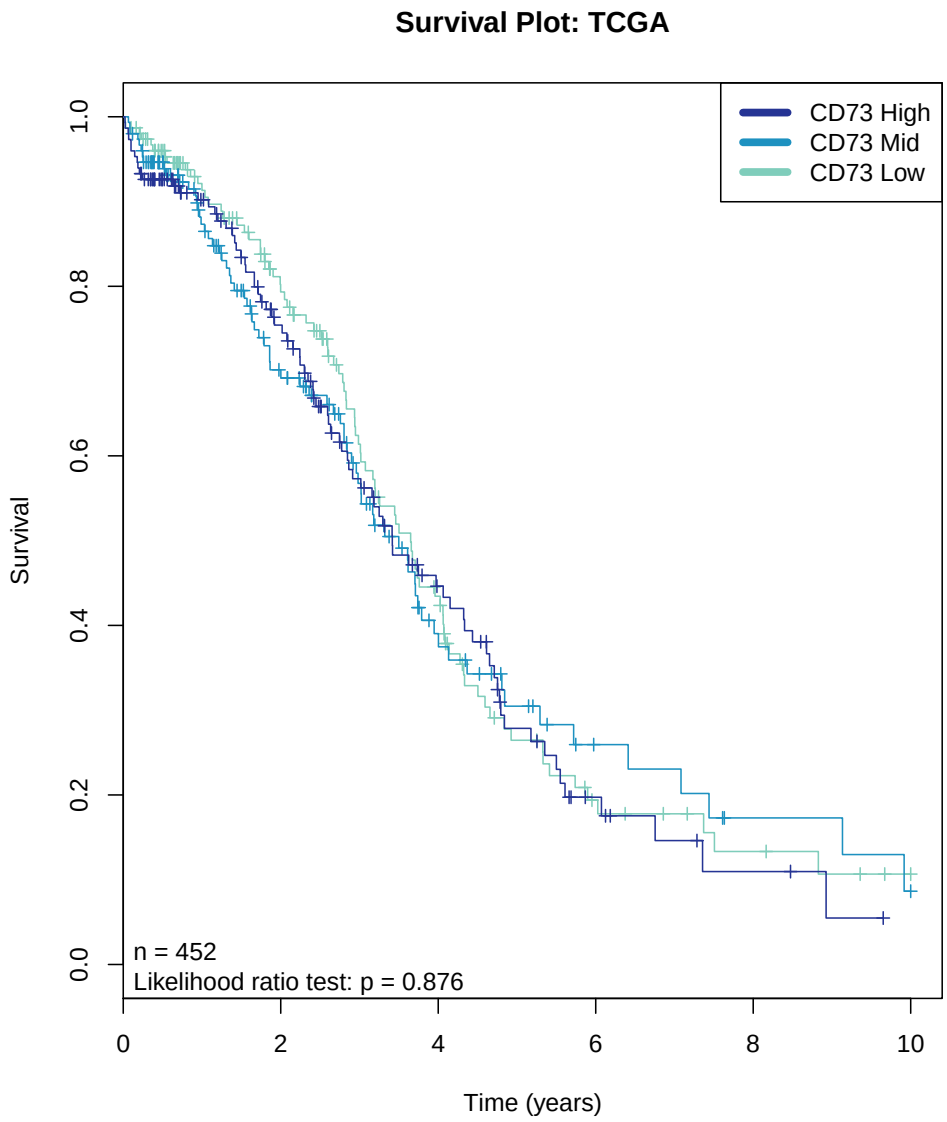


Figure 2: Survival curves for CD73 tertiles, only TCGA

Survival analysis was performed within each ovarian subtype as defined by Tothill et al., 2008 (using our implementation of their group's subtype classifier, described in Helland et al., 2011).

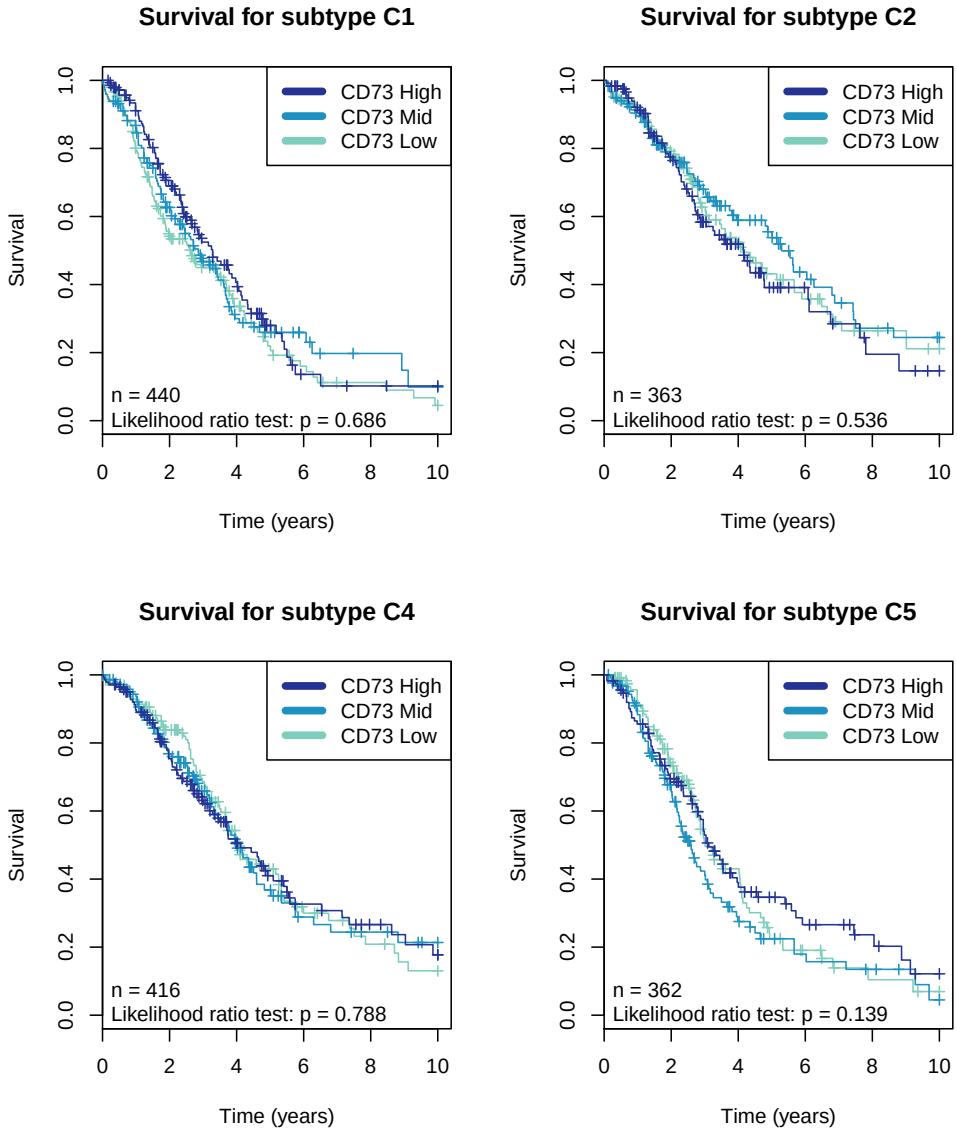


Figure 3: Survival curves for CD73 by Tothill subtype

Survival analysis was performed within each ovarian subtype as defined by TCGA, 2011 (using our implementation of the subtype classifier described in Verhaak al., 2013).

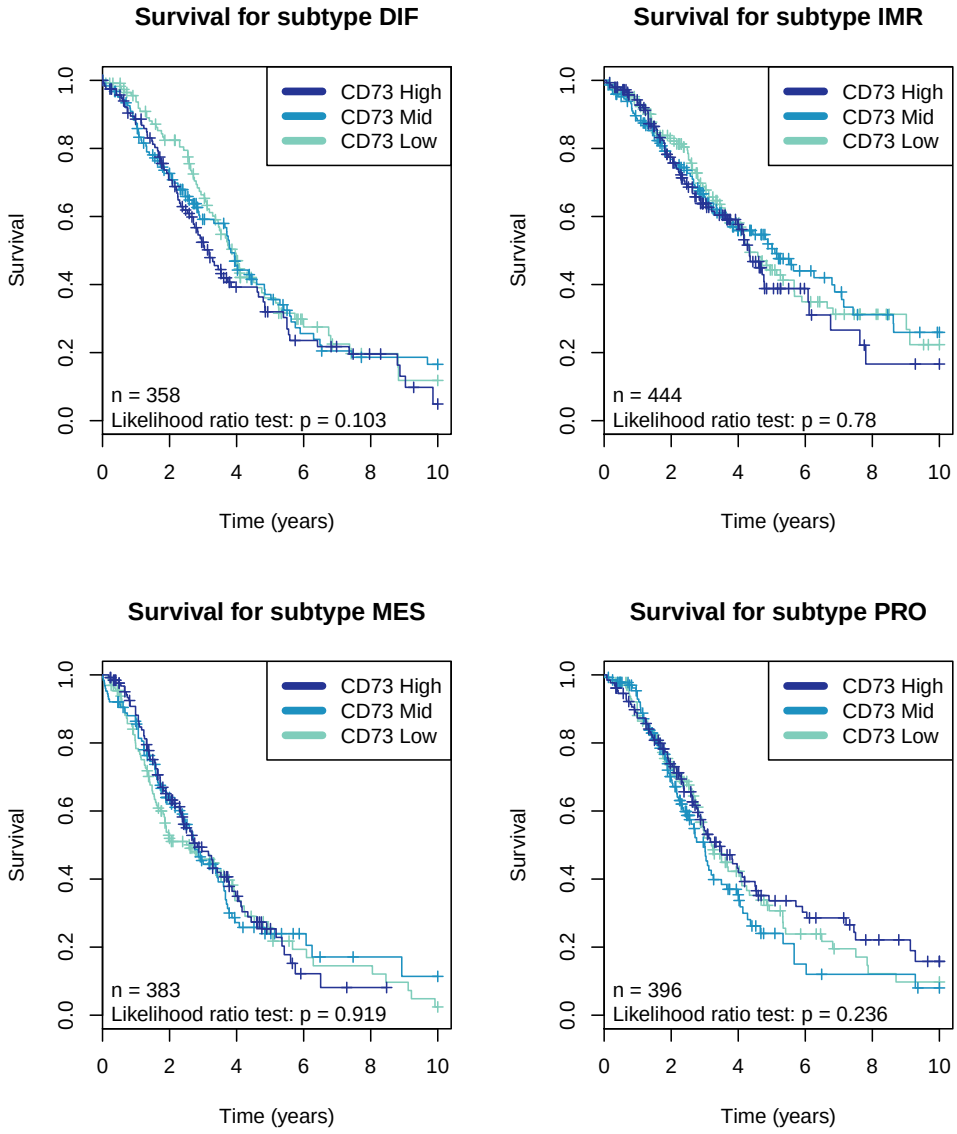


Figure 4: Survival curves for CD73 by Verhaak subtype

Patient selection config (for initial data extraction from MetaGxOvarian):

```
##------------------------------
##settings for createEsetList.R
##------------------------------

#remove retracted studies?
remove.retracted <- FALSE

#remove studies whose samples are a subset of another?
remove.subsets <- TRUE

#remove genes which had a single probe mapping to multiple genes?
## "keep":  Leave as-is, these have "///" in gene names.
## "drop":  Drop any non-uniquely mapped features.
## "split": Split non-uniquely mapped features to one per row.
##          If this creates duplicate rows for a gene, these rows are averaged.
## probes.not.mapped.uniquely <- "drop"
probe.gene.mapping <- TRUE

#rescale each gene to z-score?
rescale <- TRUE

#Keep only genes common to all platforms?
keep.common.only <- FALSE

#only keep studies with at least this many samples
min.sample.size <- 40

#only keep studies with at least this many events (deaths)
## min.number.of.events <- 15
min.number.of.events <- 0

#only keep microarray studies
min.number.of.genes <- 1000

#quantile of variance ranks above which to keep high-variance genes (ie 0.8 filters 80% of genes).
#0 means no filtering of genes
quantile.cutoff <- 0

#patient metadata which must not be missing
# meta.required <- c("days_to_death", "vital_status")

#Regexes for filtering of patients
rule.1 <- c("sample_type","^tumor$")
rule.2 <- c("histological_type","^ser$")
rule.3 <- c("summarystage","^late$")
rule.4 <- c("summarygrade","^high$")

# add Surv objects as phenoData label "y" to the esets
## add.surv.y <- function(X) Surv(X$days_to_death, X$vital_status=="deceased")

#if strict.checking is TRUE, patients missing any of the above
#metadata will be remove.  If FALSE, these patients are kept.
strict.checking <- FALSE
```

```
#if data contains missing data, should we use the impute Bioconductor package
#to impute?
## impute.missing <- FALSE

## remove which duplicates?
# FIXME
# "keep.largest" will keep duplicate in largest eset
# "keep.smallest" will keep duplicate in smallest eset
# comment out to keep all
remove.duplicates <- TRUE

#GSE19829 and GSE12418 do not have NT5E expression values for all samples. TCGA.RNASeqV2 cases are dupli
remove.datasets <- c("GSE19829", "GSE12418", "TCGA.RNASeqV2")
#remove.datasets <- c("TCGA.RNASeqV2")
```