Supplementary Figure 1. The overall copy number variation profile consisted of the called copy number gains (up, blue) and losses (down, yellow) from 7356 normal samples in arrayMap, for the whole autosomal genome content. Percentage values in Y axis corresponding to numbers of copy number imbalances, either due to regions with a high amount of germline CNVs (e.g. pericentromeric “spikes”) or based on technically challenging regions (e.g. chromosome 9 centromere, some G-C rich regions).

Supplementary Figure 2. The gene labels that are highlighted in copy number aberration profiles. These labels clearly show the locations and correlated copy number aberrations of interested genes in the chromosome.
Supplementary Figure 3. The data summary including geographic origin of liver cancer studies represented in arrayMap. In the online version, this page allows for data subset selection as well as for interactive map use.
Supplementary Figure 4. Comparing content & scope of the Progenetix and arrayMap resources.

<table>
<thead>
<tr>
<th>Techniques</th>
<th>Progenetix</th>
<th>arrayMap</th>
</tr>
</thead>
<tbody>
<tr>
<td>scope</td>
<td>tumor sample</td>
<td>array experiment</td>
</tr>
<tr>
<td>content</td>
<td>&gt;31000 samples</td>
<td>&gt;60000 arrays</td>
</tr>
<tr>
<td>raw/probe data presentation</td>
<td>no (link to sources if available)</td>
<td>yes (log2 and/or segmentation)</td>
</tr>
</tbody>
</table>
| sample re-analysis | no; supervised result (mostly as provided through publication) | yes (re-segmentation, thresholding, size filters ...)
| final data   | genome mapped CN status (interpreted) | genome mapped CN status (unsupervised) |
| main purposes | • Distribution of CNA target regions in most tumor types (>350 ICD-O) • Cancer classification | • Gene specific hits • Genome feature correlation (fragile sites ...) |
Supplementary information - API: R

This is an example scenario for the arrayMap/Progenetix API in an R environment, making use of the “pgDataLoader.R” library accessible through https://github.com/progenetix/pgRpi/. Here, the survival correlation of the MYCN CNA status is shown for medulloblastoma samples (ICD-O 9500/3).

```r
rm(list = ls())
library(survival)
source("~/lib/Rlibs/pgDataLoader.R")
icdm <- c("9500/3")
gene <- "MYCN"
survData <- pgDataLoader(pg.server="127.0.0.1", api_out="samples", icdm_m=icdm, db="arraymap", genes_m=paste(gene, ",", sep=""))
nrow(survData)
geneColumn <- paste('GENE_', gene, sep="")
genePos <- grep(geneColumn, colnames(survData), perl="T")
gainNo <- nrow(subset(survData, survData[genePos] == 1))
plot(survfit(Surv(survData$FOLLOWUP, survData$DEATH) ~ survData[genePos] == 1, se.fit=TRUE), col=c("black","blue"), main=paste("Survival and gene ", gene, " (ICD-O ", icdm, ",")", sep=""), xlab="months", ylab="survival", cex=1.2)
sdf <- survdiff(Surv(survData$FOLLOWUP, survData$DEATH) ~ survData[genePos] == 1)
pcsq <- round(pchisq(sdf$chisq, df=1, lower=FALSE), digits=5)
legend("bottomright", c("no gain", paste(gene, ' gain (', gainNo, '/', nrow(survData), ')', sep="")), fill=c("black","blue"), inset=c(0.02,0.02), bg="azure1", cex=0.8)
legend("bottomleft", c(paste("p:", pcsq)), inset=c(0.02,0.02), bty="n", cex=1)
```

**Survival and gene MYCN (ICD-O 9500/3)**

![Survival and gene MYCN (ICD-O 9500/3)](image-url)