CAIRO User Story 2

User	Mariska Bierkens, Rinus Voorham, Linda Bosch
CTMM Project(s)	DeCoDe
Role in project(s)	Researcher
Goal	To validate MSI as a prognostic biomarker for disease progression in metastatic colorectal cancer
Benefit	Combine experimental and clinical data to perform analyses (survival- and frequency plots), and to make data publicly available

Steps to answer research question (current situation)

1. Acquire clinical data (Excel file)
2. Tumor classification (TNM staging, MSI high/low) by histopathology
3. Acquire genomics data, for CNV (aCGH) on 2 levels:
   -- call data: gain, loss, neutral, amplification per segment mapped back onto the probes
4. Define regions: chromosome start stop locations and region name <-- 'Gene List' / 'Region List'
5. Molecular analysis: for same locus (or even different locations) : loss + methylation, gain + non methylation etc : 6 cell table with methylated / non methylated vs loss, neutral, gain - 'CGH test'
6. Create survival analysis KM plot with different arms with R or SPSS:
   --> aCGH call data (loss, gain etc on a specific locus, in a gene or even on user defined chromosomal region)
   --> methylation data (yes/no)

Desired situation

1. Clinical data and aCGH call data and methylation yes/no data and mutations data (KRAS, BRAF) and MSI data all available
2. Be able to define chromosomal regions by chromosome, start/stop, genome build and give them a name (e.g. 20q gain)
3. **Review the CNV profiles of each sample that were created in the QC steps of the aCGH call pipeline (WP4)**
4. (evt. Goos/Belt: review the TMA scores for a certain protein by drawing a boxplot of all the scores; be able to click an outlier and see the underlying core intensity and frequency scores and the scanned images of the cores or even the original slides --> tEPIS --> WP2: can tEPIS import images from Pannoramic Viewer?)
5. **Be able to select chromosomal region or gene and build 4x2 cross table with for example aCGH call (loss, neutral, gain, amplification) vs methylation yes/no and show appropriate test statistics**
6. (Be able to select chromosomal region or gene, define two (patient) groups, and show boxplots of DNA copy number rations, click on outliers to view gene region)
7. **Genome wide analysis --> 2 groups: which chromosomal regions differ significantly on mutation data / clinical data / --> view in list with p-value (Call CGHTest in background --> Multiple testing correction)**
8. Nice to have: R script standardization: your favourite data visualization / statistics method
9. Combine clinical and biomarker data to perform survival analyses and frequency plots – treatment arms should be defined either on clinical or molecular data
   --> view plot, P value (which test? logrank, cox etc), and # events and HR

Dataset

A large study on genomic profiling and survival times, in colorectal cancer patients. This was done by classification of MSI (low or high MicroSatellite Instability),) and acquiring genomic data, where CNV is measured using aCGH.
Various other clinical variables and biomarkers were measured, such as treatment and surgery information, tumor classification, survival time in months.

Cohort information

CAIRO 1 study: 820 colorectal cancer patients in 74 Dutch hospitals, in 24 months.

Clinical data

The clinical data is available in Excel, but is also being stored into the TraIT OpenClinica instance.

Biobanking

Tumor tissue samples.

Experiment data

Tumor samples were classified for low or high MSI, and genomics data was acquired CNV calculation. This was done using aCGH segmentation and calling.

Research Question

What is the prognostic value of genetic profiles, low and high MSI, in CRC liver metastases and their corresponding primary tumors?

Example workflow files

Excel file containing clinical data

CAIROnr	gender	treatment arm	Date of randomisation	Age	Best response according to RECIST 1st line	PFS1	PFS1event	Overal survival	Cause of death	Site primary tumor	Location metastases	MSI		
1	F	Sequential (A)	1.1.1900	58	PR	60								
2	F	Combination (B)	1.1.1900	73	PR	120								
3	M	Combination (B)	1.1.1900	60	early death toxicity	548								

Input R-script: Call data per probe, tumor versus normal tissue (tab-delimited)

Probe name	Chromosome	Start	Cairo.12.T.vs.N	Cairo.13.T.vs.N	Cairo.14.T.vs.N	Cairo.15.T.vs.N	Cairo.16.T.vs.N	etc						
A_16_P00000027	1	784458	-1	0	0	-1	-1							
A_16_P00000036	1	799631	-1	0	0	-1	-1							
A_16_P00000037	1	802868	-1	0	0	-1	-1							
etc														

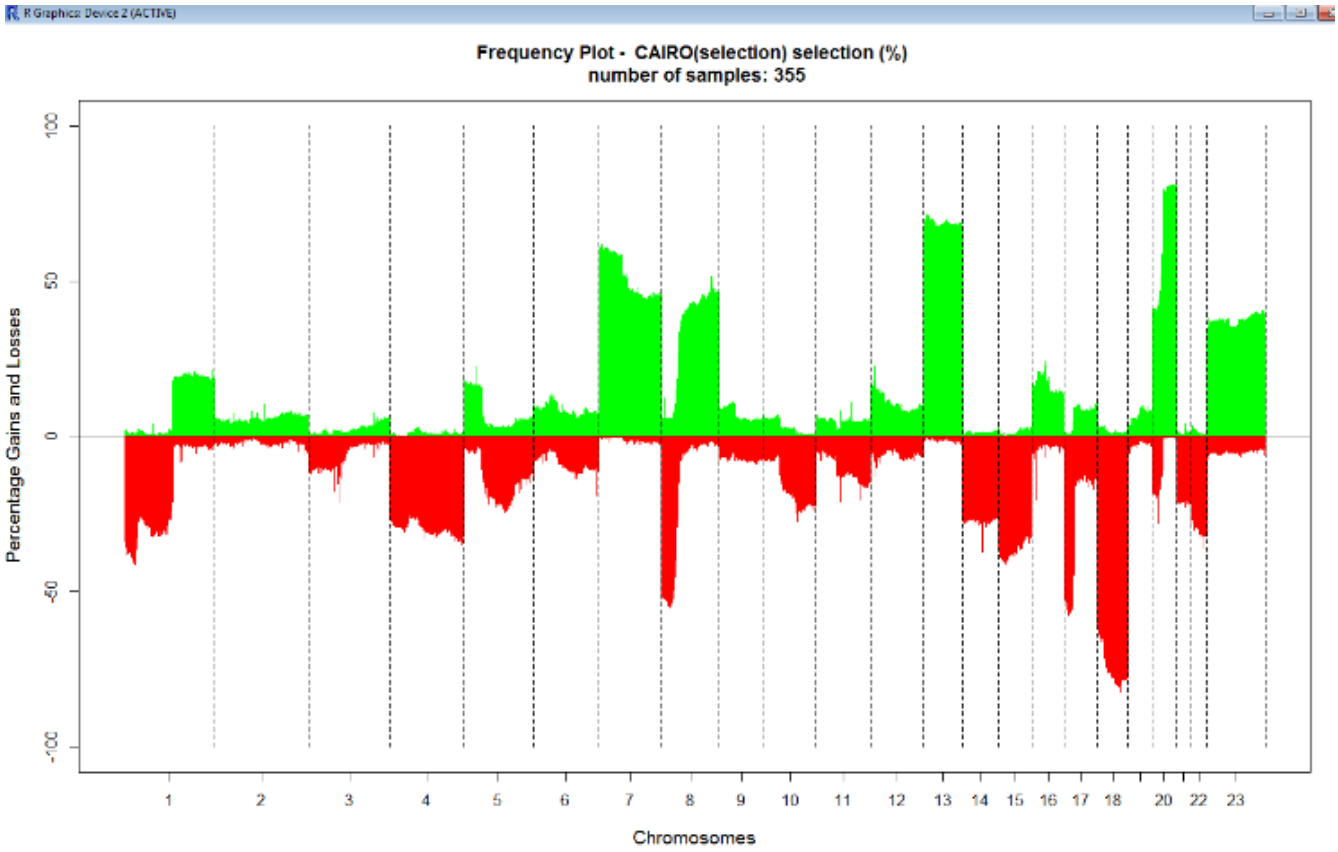
Input R-script: values per segment

Sample	values	start	end	chrs	nclone					
Cairo.2.T.vs.N	-0.211	1	10757	1	10757					
Cairo.2.T.vs.N	0.177	10758	10762	1	5					
Cairo.2.T.vs.N	-0.212	10763	13275	1	2513					
Cairo.2.T.vs.N	0.205	13276	13281	1	6					
Cairo.2.T.vs.N	-0.191	13282	13714	1	433					
Cairo.2.T.vs.N	-0.003	13715	28028	2	14314					
etc										

Output R-script: segmented CNV values, mapped back on probes

Probe name	Chromosome	Start	Cairo.2.T.vs.N	Cairo.3.T.vs.N	Cairo.4.T.vs.N	Cairo.5.T.vs.N	Cairo.6.T.vs.N	etc						
A_16_P00000027	1	784458	-0.211	0.003	0.041									
A_16_P00000036	1	799631	-0.211	0.003	0.041									
A_16_P00000037	1	802868	-0.211	0.003	0.041									
etc														

Output R: Frequency plot, gains losses per chromosome


R Graphics: Device 2 (ACTIVE)
Frequency Plot - CAIRO(selection) selection (%)
number of samples: 355
Percentage Gains and Losses
100
50
0
-50
-100
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
20
22
23
Chromosomes