CLL pilot - Rainfall analysis

Dr. Susanne Weller

September 29, 2014

Contents

Т	Data preparation	1
2	Number of Kataegis events	7

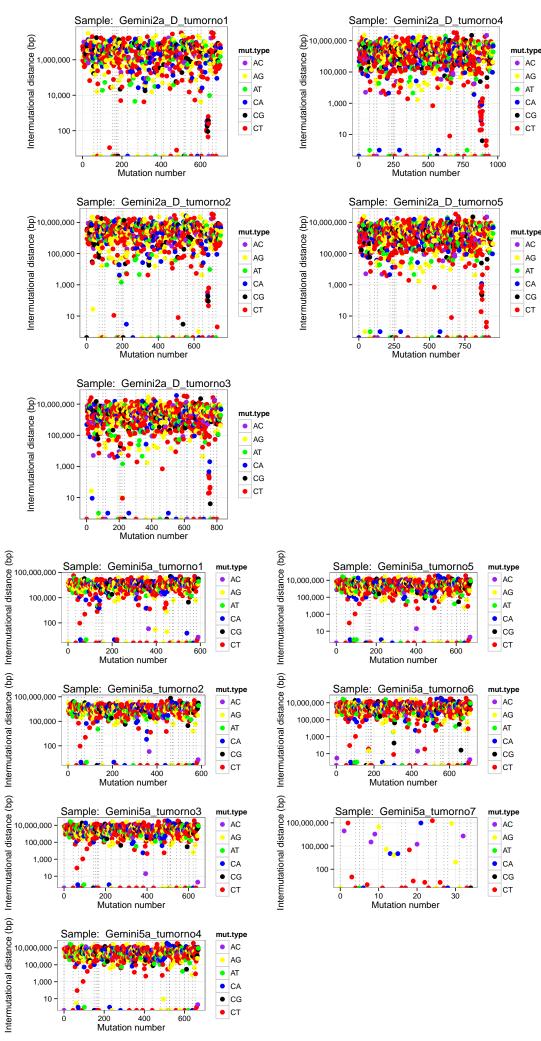
1 Data preparation

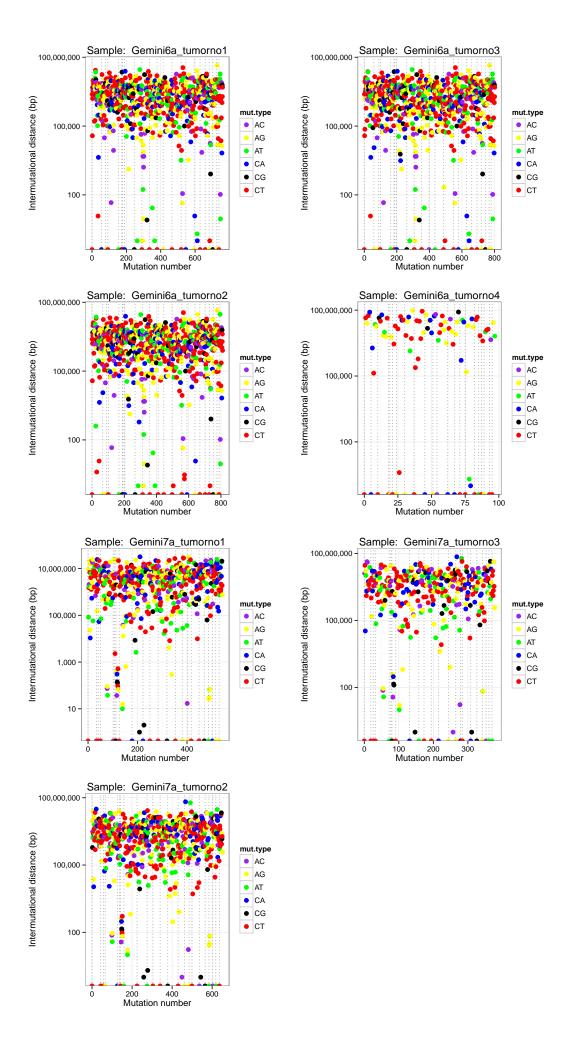
Make sure that you have loaded the following R packages: ${\tt ggplot2}$ ${\tt reshape2}$

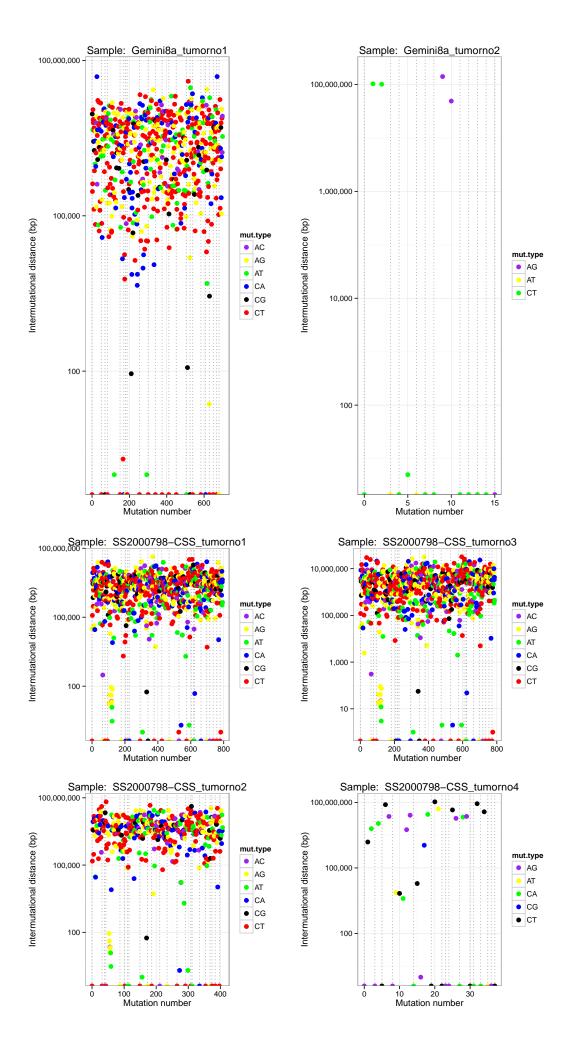
data.table

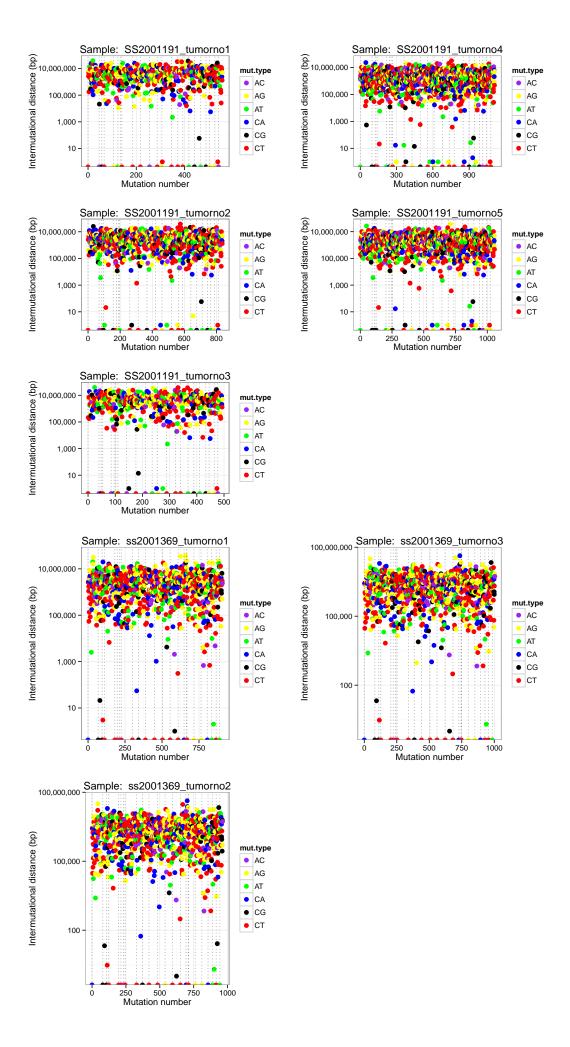
SNV data were obtained from Illumina amd augmented with tumor numbers for sequential tumors and only snvs selected for the analysis. A python script is used to calculate intermutational distances.

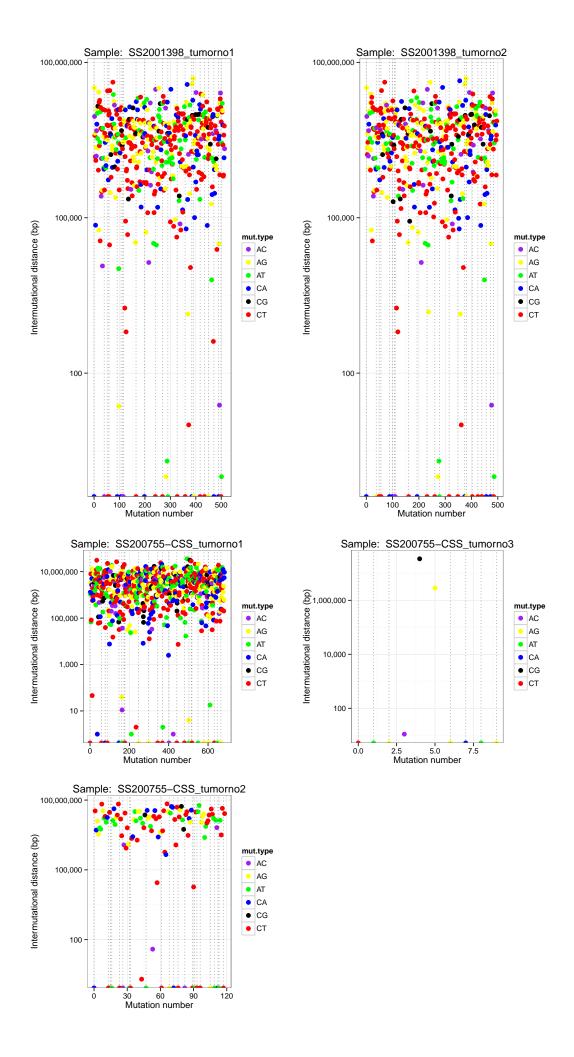
To create Kataegis plots, intermutational distances were calculated within each chromosome and plotted against chromosome numbers. Please play around with width and hight in R Sweave or export directly to pdf for appropriate sizes.











2 Number of Kataegis events

From Lawrence 2013, nature:

"Kataegis events, detected by identifying stretches of at least six mutations having inter-mutation distances at least two standard deviations smaller than the sample median, are indicated with blue boxes at the bottom of the plots. In some cases it can be observed that the individual mutations of a kataegis event are all of the same category."

Because we have a very low number of kataegis events anyway, I counted each event separately, using the 10% quantile as a cutoff. Note that the low number of kataegis events in some of the final samples might come from germ line contaminations.

