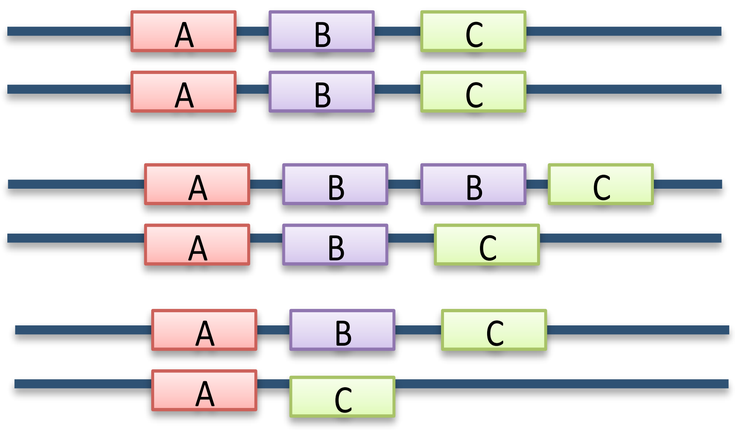
Gene-based analysis of focal copy number aberration patterns in cancer genomes - A step closer to understanding potential driver functionality?

Saumya Gupta and Michael Baudis
Institute of Molecular Life Sciences, University of Zürich, Switzerland
Swiss Institute of Bioinformatics, University of Zürich, Switzerland

CNAs in Cancer Genomes

Cancers are genomic disease, characterized by the accumulation of :
Point mutations (insertions, deletions, substitutions)
Chromosomal rearrangements
Regional Copy Number Alterations (losses, gains)
Epigenomic changes (e.g. DNA methylation abnormalities)

On average, more than 14% of a given cancer genome is in an imbalanced state due to gain/loss of regions (from several kb to whole chromosomes), potentially affecting many genes.

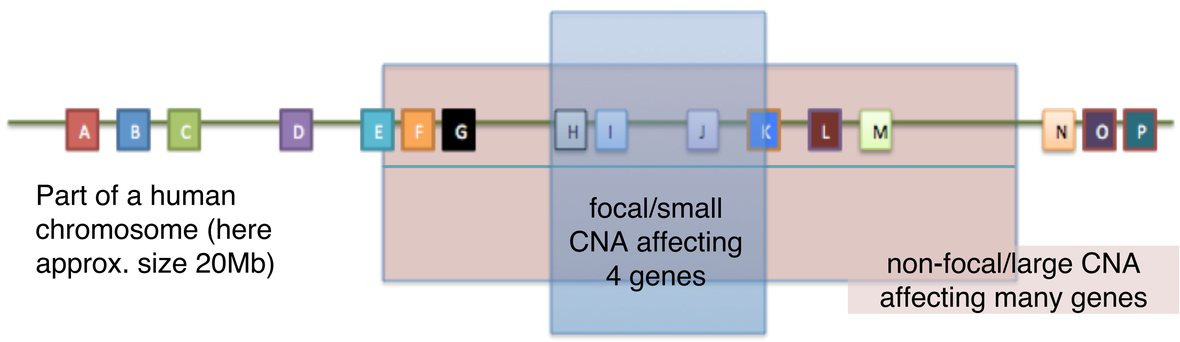


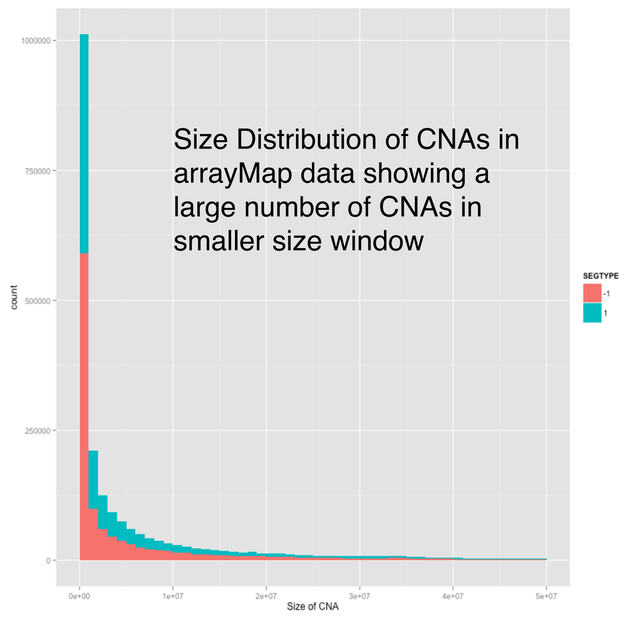
Normal individual has 2 copies of each gene

3 copies of Gene B amplification (+1)

Only 1 copy of Gene B deletion (-1)

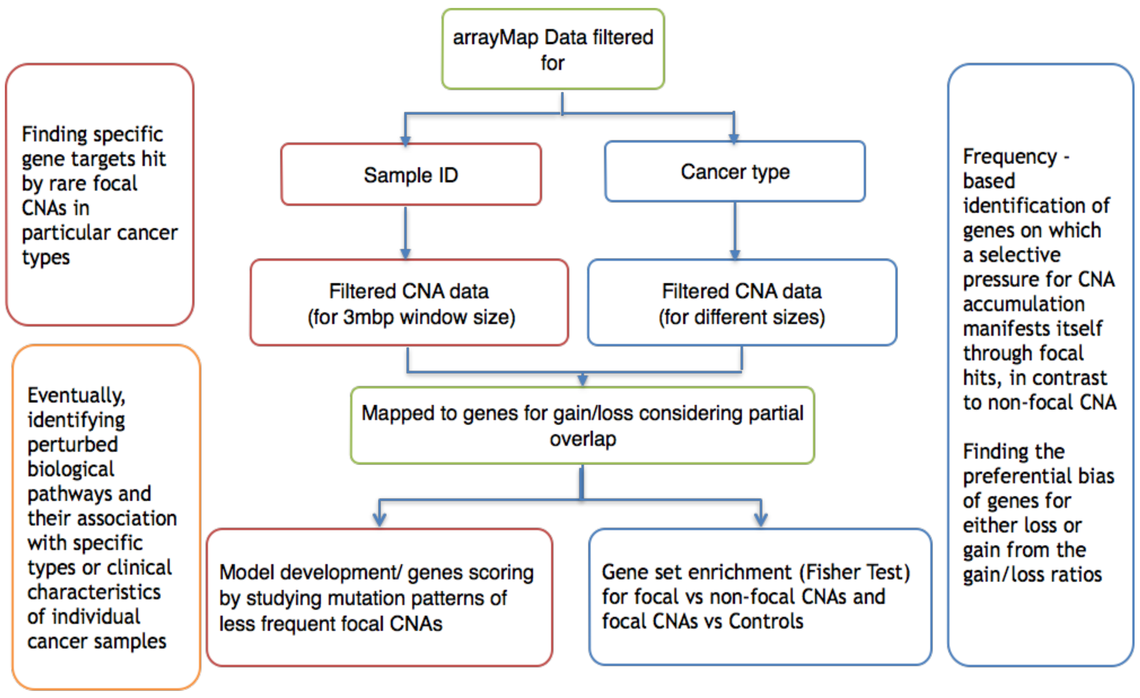
Need to study focal CNAs



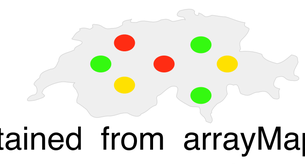


Since many CNAs involve larger chromosomal regions and affect many genes, it becomes difficult to identify genes (from a group of genes) that may have cancer specific role. With increasing size of a CNA, the probability to hit unspecific “bystander” targets increases.
In this study, focal CNAs are defined (operationally) as 3 Mbp or smaller in size[1]
Determining the "focality" of CNAs precisely targeting a gene could give information about the potential driver functionality of genes mapping to the regions affected

Methodology & Objectives



Data Sources

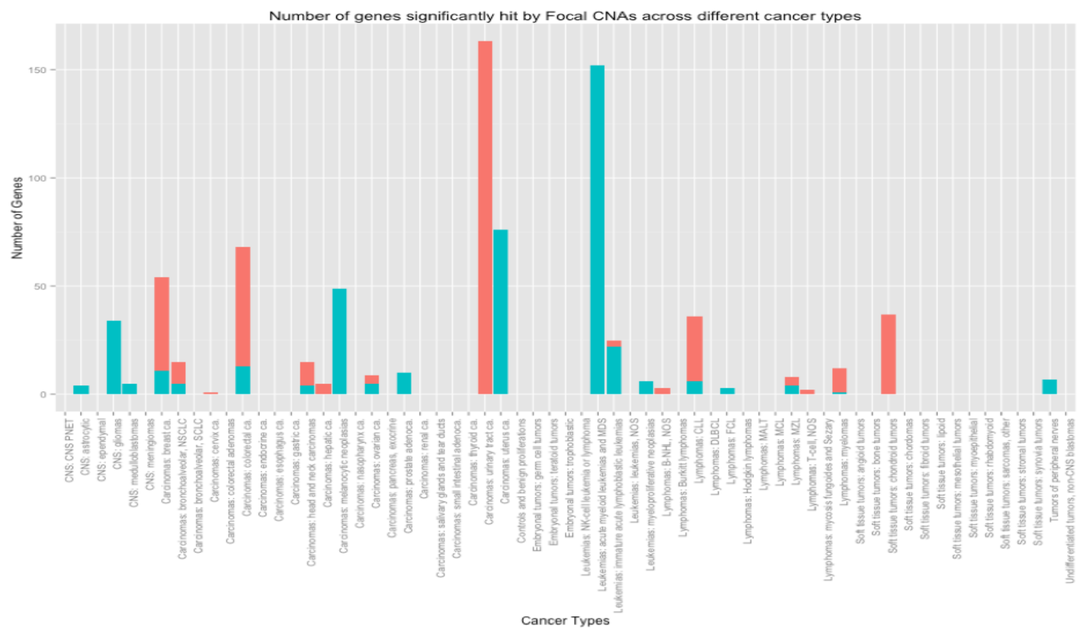


arrayMap[2]: The CNA data for all the cancers was obtained from arrayMap database (http://www.arraymap.org/cgibin/amHome.cgi). The current version of data reflects 65,075 genomic copy number arrays collected and processed from NCBI GEO, EBI ArrayExpress, The Cancer Genome Atlas, publication supplements and user submitted data.

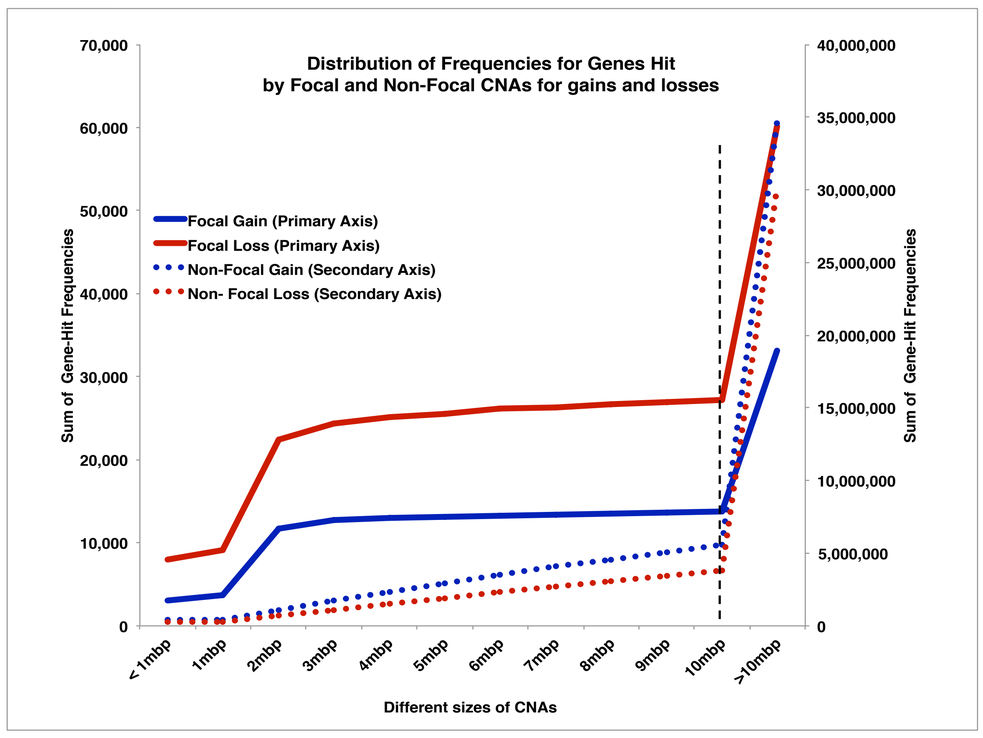
Gene-hit frequencies across different sizes of CNAs

Gain												
Genes	**900kbp**	**1mbp**	**2mbp**	**3mbp**	**4mbp**	**5mbp**	**6mbp**	**7mbp**	**8mbp**	**9mbp**	**10mbp**	**>10mbp**
OR4N4	372	446	1,608	1,718	1,745	1,760	1,771	1,781	1,793	1,810	1,821	2,320
POTEC	222	350	1,574	1,670	1,691	1,704	1,712	1,719	1,724	1,737	1,744	2,172
FAM53A	45	49	174	225	284	316	336	365	405	520	1,312	1,823
SLBP	44	48	171	222	281	313	333	362	402	517	1,309	1,820
Loss												
OR4K2	414	440	902	943	969	996	1026	1038	1045	1057	1062	1786
TP53TG3	23	36	234	294	322	325	336	342	343	344	344	1421
ABR	188	199	373	468	533	575	615	689	769	846	959	3681
FAM90A14	476	501	584	613	634	671	699	737	810	901	959	4519

Gene set enrichment for genes hit by focal CNAs



In this graph, the number of genes amplified or deleted significantly in different cancer types for focal CNA . For focal CNAs, some cancer types show a “preference” for either gain or loss of genes.



The distribution of somatic CNAs in cancer genomes. In this graph blue lines denote gains and red lines denote losses. Different trends of accumulative hit frequency for focal and non-focal CNAs are denoted by solid and dotted lines respectively.

Conclusions

Focal CNAs that are limited to only one or few genes could point to cancer drivers even if the CNAs are not in copy number hotspots

Future analysis will focus on genes hit by focal CNAs with preference to either gain or loss to reach to more specific targets

References

[1] Candidate driver genes in focal chromosomal aberrations of stage II colon cancer, Brosens et. al., Journal of Pathology, 2010

[2] arrayMap: A Reference Resource for Genomic Copy Number Imbalances in Human Malignancies. Haoyang Cai, Nitin Kumar, Michael Baudis PLoS One. 2012