Genome-Wide Co-Localization of Somatic Copy Number Alterations and Germline Common Variant Risk Loci in Cancer

Marcin Imielinski M.D, Ph.D.

November 17th, 2011

Dept of Pathology, Brigham and Women's Hospital

Meyerson Lab, Broad Institute / Dana Farber Cancer Institute

Two facets of cancer genomics

- Germline risk variants
  - inborn mutation / variation that affects lifetime cancer risk
- Somatic alterations
  - develop in the tumor and directly contribute to tumorigenesis, metastasis, drug resistance, et al.

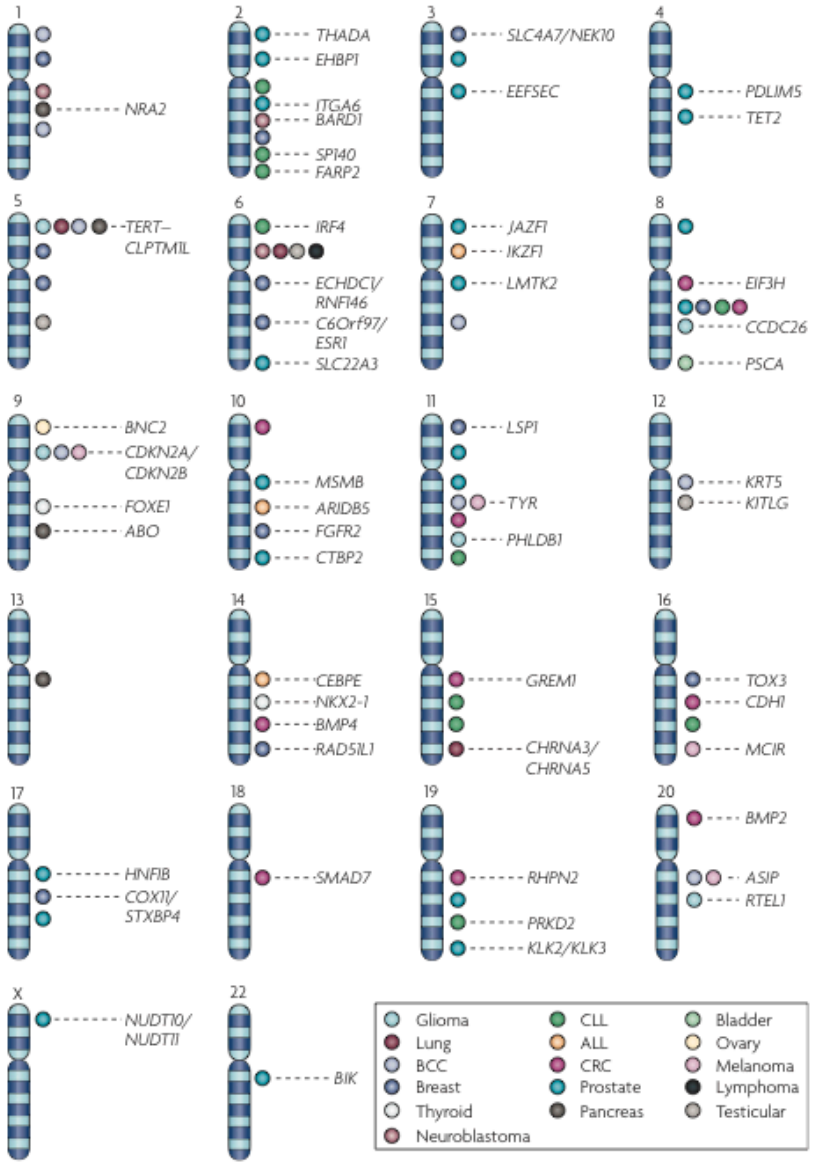
Heritable Cancer Risk

- Common cancers are 2-4 fold more likely in first degree family members of affected pts
- Heritability varies from 0 to 0.40 depending on tumor type, gender, age of onset et al. (across total cancer 0.07-0.25[1])
- Cancer risk mediated by complex polygenic inheritance

[1]Ahlbom et al JNCI 1997

Heritable Cancer Risk

- Rare, highly-penetrant variants
  - *TP53* (Li-Fraumeni), *APC* (FAP), *MLH / MSH* (HNPCC), *BRCA1 / BRCA2* (Familial Breast cancer), *RB1* (Retinoblastoma)
  - Explain 5% and 20% of heritable risk for breast and CRC, respectively.
- Common, mildly-penetrant variants
  - Last ~5 years of GWAS ~ 300 loci in 20 cancer types
  - Explain 10%, 23%, and 6% of heritable risk for breast, prostate, and CRC, respectively



Common variant cancer susceptibility loci (circa early 2010)

Fletcher and Houlston

Germline x Soma

- Examples of germline cancer susceptibility loci that are frequently mutated in cancer ...
  - *TP53*
  - *APC*
  - *RB1*
  - *CDKN2A*
  - *VHL*
  - *NF1*

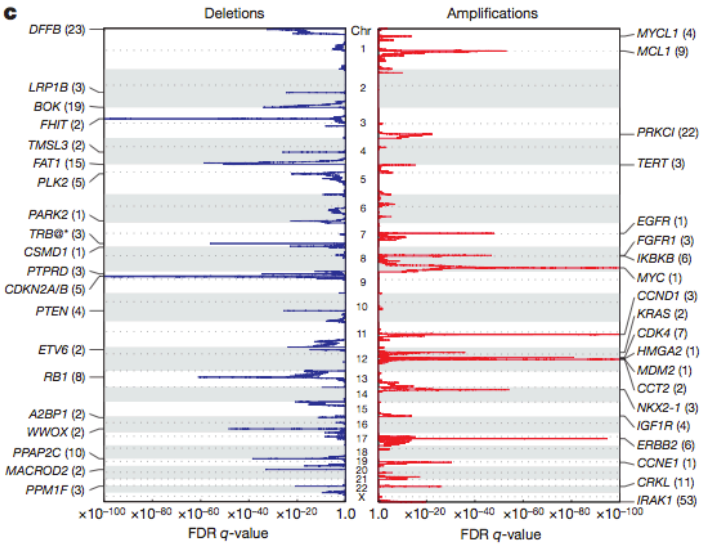
GWAS x SCNA

- Genome wide data:
  - Germline: 297 cancer loci from 85 GWASs
  - Somatic: 258 SCNA peak regions from Global Cancer Map (GCM) study (Beroukhim, Mermel et al *Nature* 2010)
- Approach:
  - Quantify overlap, determine significance against null model built via permutation

Vol 463|18 February 2010|doi:10.1038/nature08822 nature

ARTICLES

The landscape of somatic copy-number alteration across human cancers



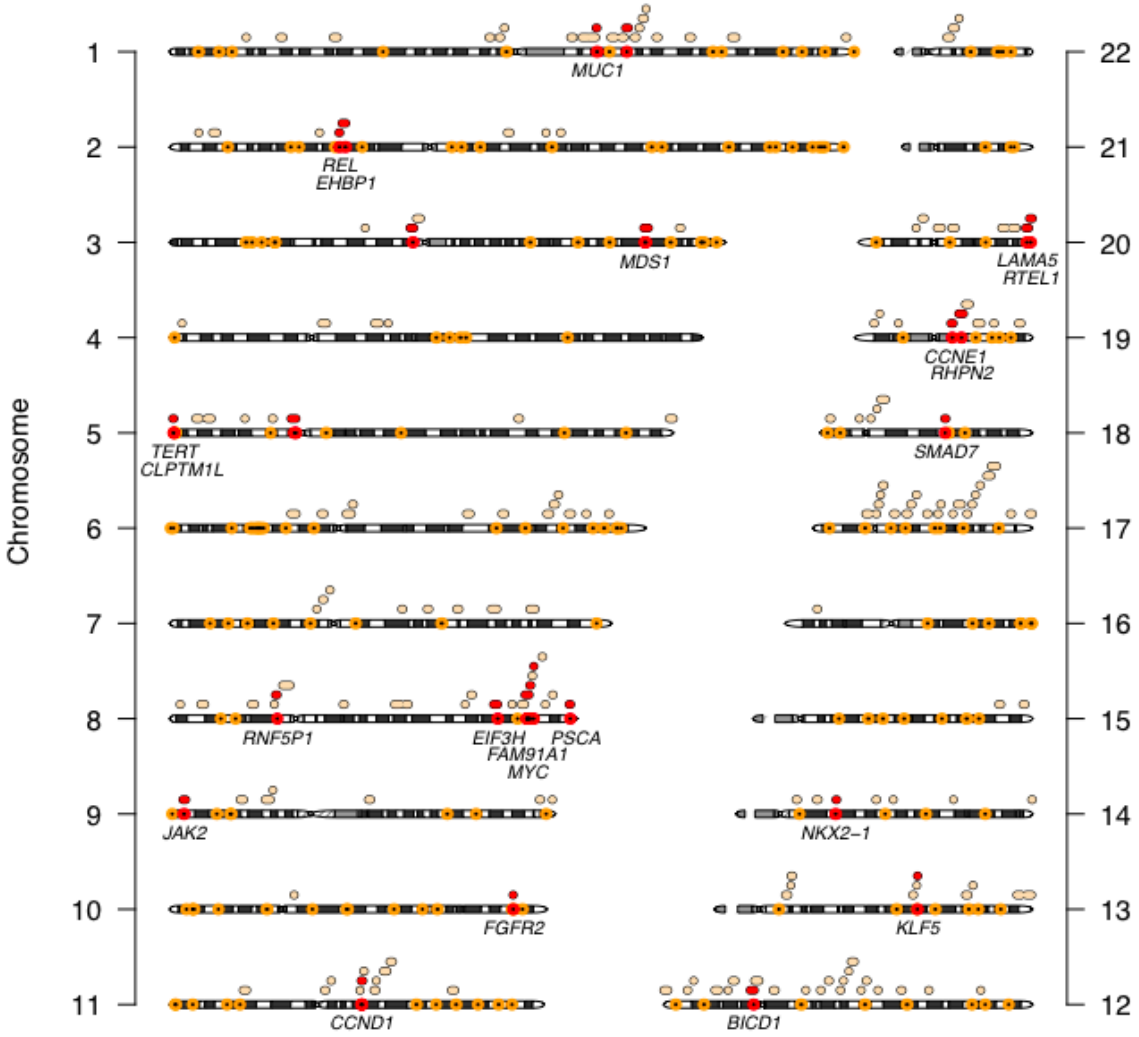
Data: GWAS

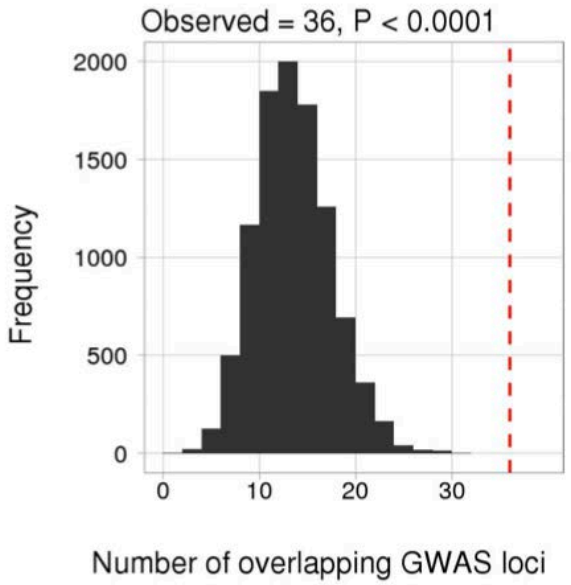
- 297 Loci from NHGRI GWAS database
  - Report most significant SNPs reported by published GWASs
  - Hand-picked “cancer GWAS’s”
    - 20 traits
  - For each SNP defined “locus” as LD neighborhood with r^2 > 0.3 in CEU HapMap
- → 219 unique GWAS loci (1.1% of the mappable genome)

Data: SCNA

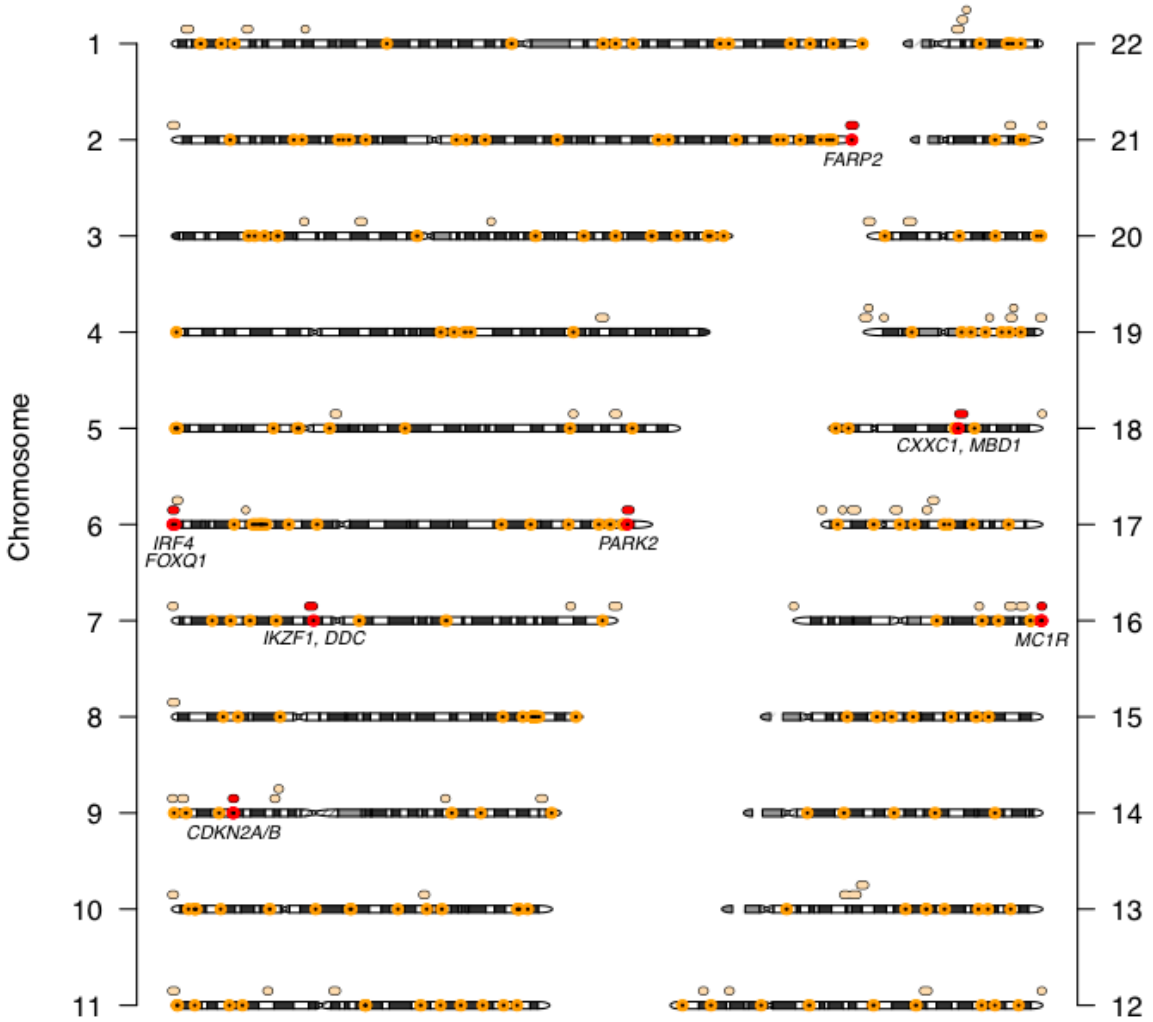
- Significant (q<0.25) SCNA peak regions from GCM pan tumor and 20 tumor subanalyses
  - pan tumor analysis, epithelial, hematopoietic, and 18 other specific tumor types
- Combined all amps, dels, and all peaks achieving significance in *any* tumor type sub-analysis
  - 198 amp hotspots (5.8% of mappable genome)
  - 67 del hotspots (2.6% of mappable genome)
  - 258 total SCNA peak regions (8.4% of mappable genome)

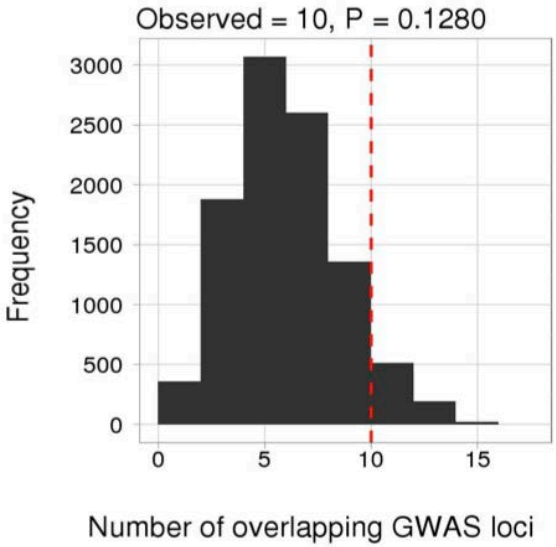
GWAS loci vs Amplification peak regions





GWAS loci vs Deletion peak regions





Analysis of cancer vs non-cancer associated GWAS LD regions

Amplification SCNA

	Non-cancer GWAS locus	Cancer GWAS locus
Does not intersect amplification peak	2235	183
Intersects amplification peak	179	36

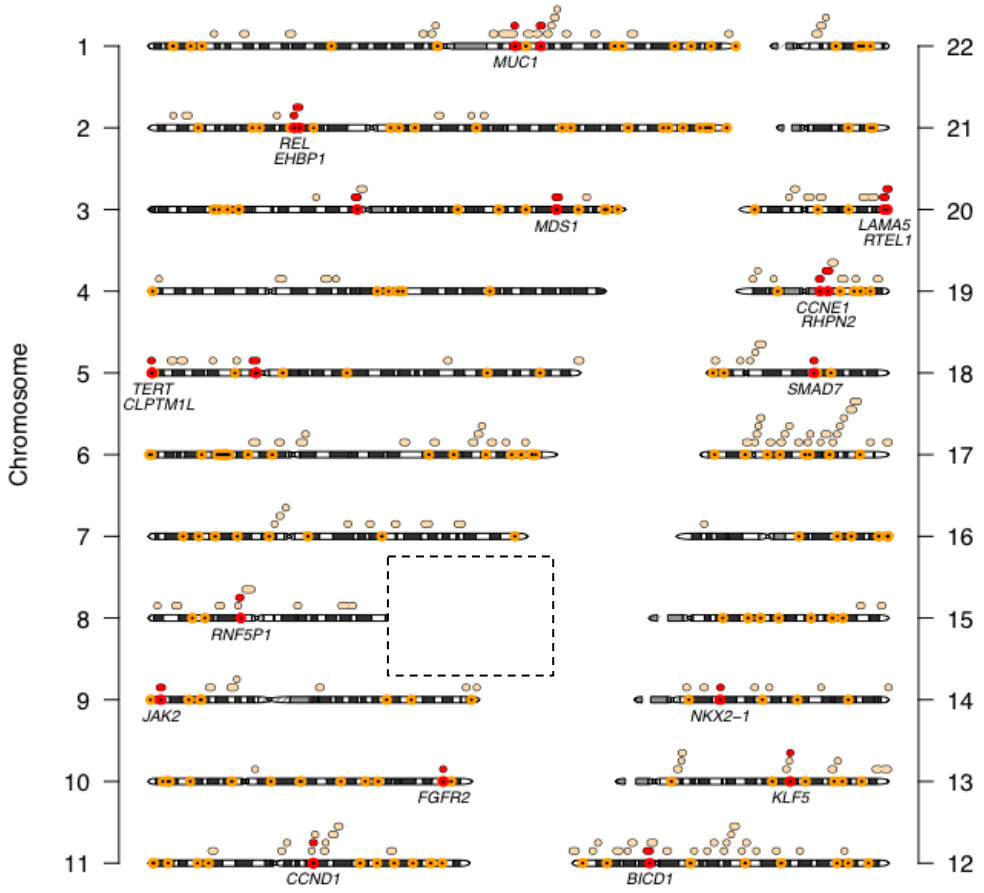
$P = 2.5 \times 10^{-5}$,
Fisher's exact test

Deletion SCNA

	Non-cancer GWAS locus	Cancer GWAS locus
Does not intersect deletion peak	2336	209
Intersects deletion peak	78	10

$P = 0.32$,
Fisher's exact test

Analysis of cancer vs non-cancer associated GWAS LD regions (after removing MYC locus)



Amplification SCNA

	Non-cancer GWAS locus	Cancer GWAS locus
Does not intersect amplification peak	2235	183
Intersects amplification peak	174	26

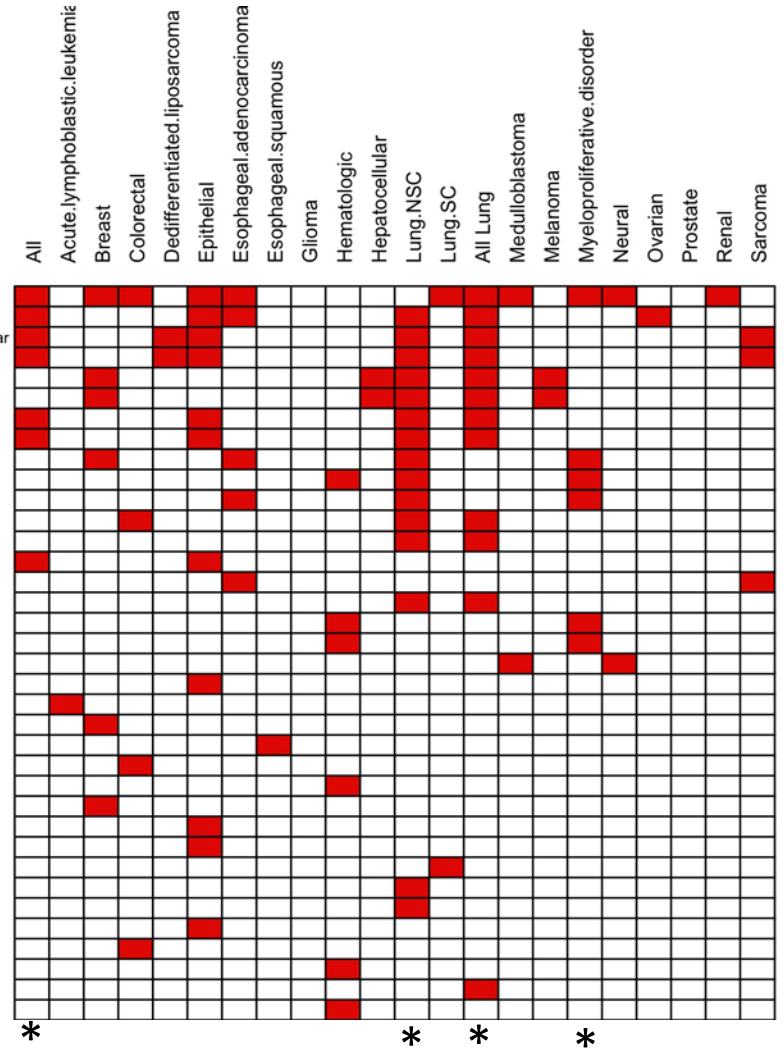
P = 0.0096,
Fisher's exact test

GWAS loci x Tumor Type-Specific SCNA peaks

GWAS Loci

Reported genes	Region	Traits
MYC	chr8:128.782-128.787	Bladder
CCNE1	chr19:34.933-35.037	Bladder
TERT	chr5:1.34-1.346	Glioma, Lung adenocarcinoma, Testicular
CLPTM1L	chr5:1.347-1.409	Lung adenocarcinoma
Intergenic	chr11:68.941-69.001	Renal cell carcinoma
MYEOV, CCND1, et al	chr11:69.017-69.055	Breast
NKX2-1	chr14:35.549-35.78	Thyroid
RTEL1	chr20:61.666-61.878	Glioma
Intergenic	chr8:128.544-128.609	Prostate
Intergenic	chr8:128.35-128.465	Breast
ORF DQ515897	chr8:128.476-128.525	Colorectal
PVT1	chr8:129.225-129.294	Hodgkin's lymphoma
Intergenic	chr11:68.71-68.794	Prostate
BICD1	chr12:32.259-32.343	Pancreatic
RHPN2	chr19:38.219-38.363	Colorectal
EHBP1	chr2:62.881-63.259	Prostate
Intergenic	chr8:128.145-128.207	Prostate
Intergenic	chr8:128.245-128.282	CLL
PSCA	chr8:143.749-143.854	Bladder
NR	chr1:153.335-153.585	Esophageal and gastric
Intergenic	chr1:164.098-164.169	Testicular tumor
FGFR2	chr10:123.323-123.345	Breast
KLF5, KLF12	chr13:72.777-72.831	Pancreatic
SMAD7	chr18:44.703-44.713	Colorectal
REL	chr2:60.895-61.018	Hodgkin's lymphoma
LAMA5	chr20:60.324-60.417	Colorectal
MDS1,EVI1	chr3:170.525-170.587	Nasopharyngeal carcinoma
MYNN	chr3:170.899-171.065	Colorectal
Intergenic	chr3:87.172-87.453	Prostate
Intergenic	chr5:44.612-45.11	Breast
Intergenic	chr5:45.126-45.322	Breast
EIF3H	chr8:117.666-117.884	Colorectal
MYC,THEM75	chr8:129.562-129.662	Ovarian
CCDC26	chr8:130.522-130.755	Glioma
RNF5P1	chr8:38.573-38.606	Pancreatic
JAK2	chr9:4.972-5.261	Myeloproliferative neoplasms

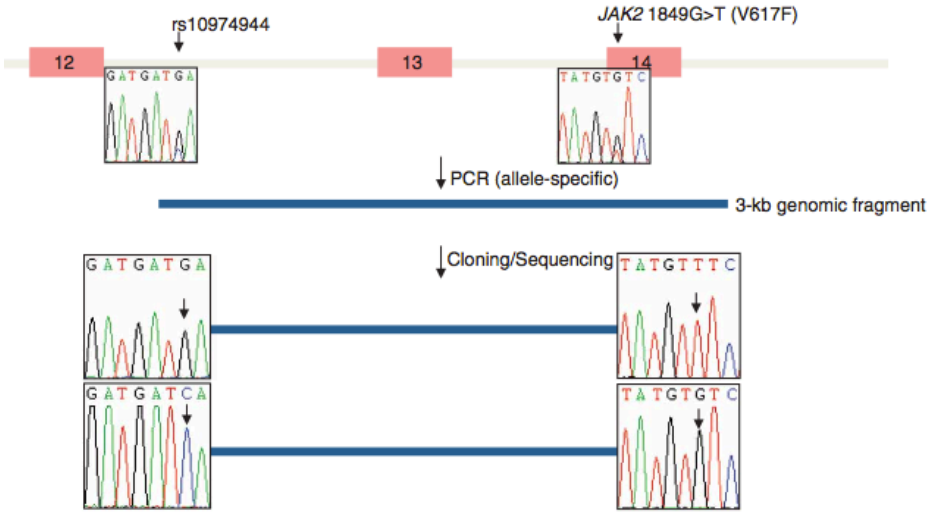
SCNA tumor type analyses



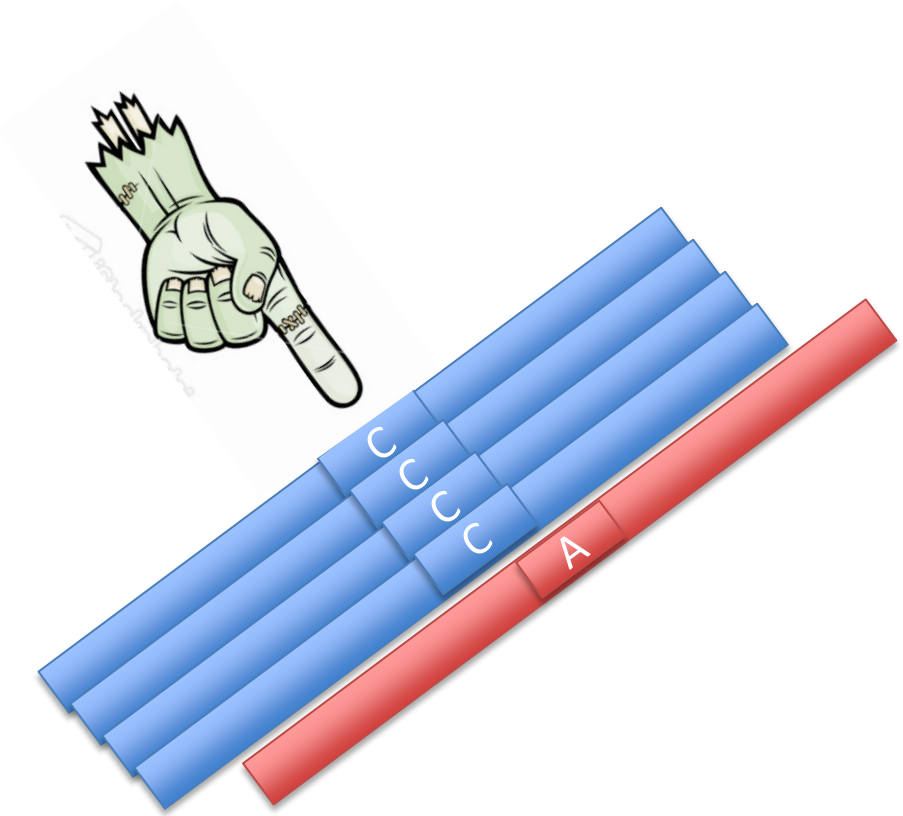
Does germline SNP status confer risk for specific somatic alterations?

A germline *JAK2* SNP is associated with predisposition to the development of *JAK2*^{V617F}-positive myeloproliferative neoplasms

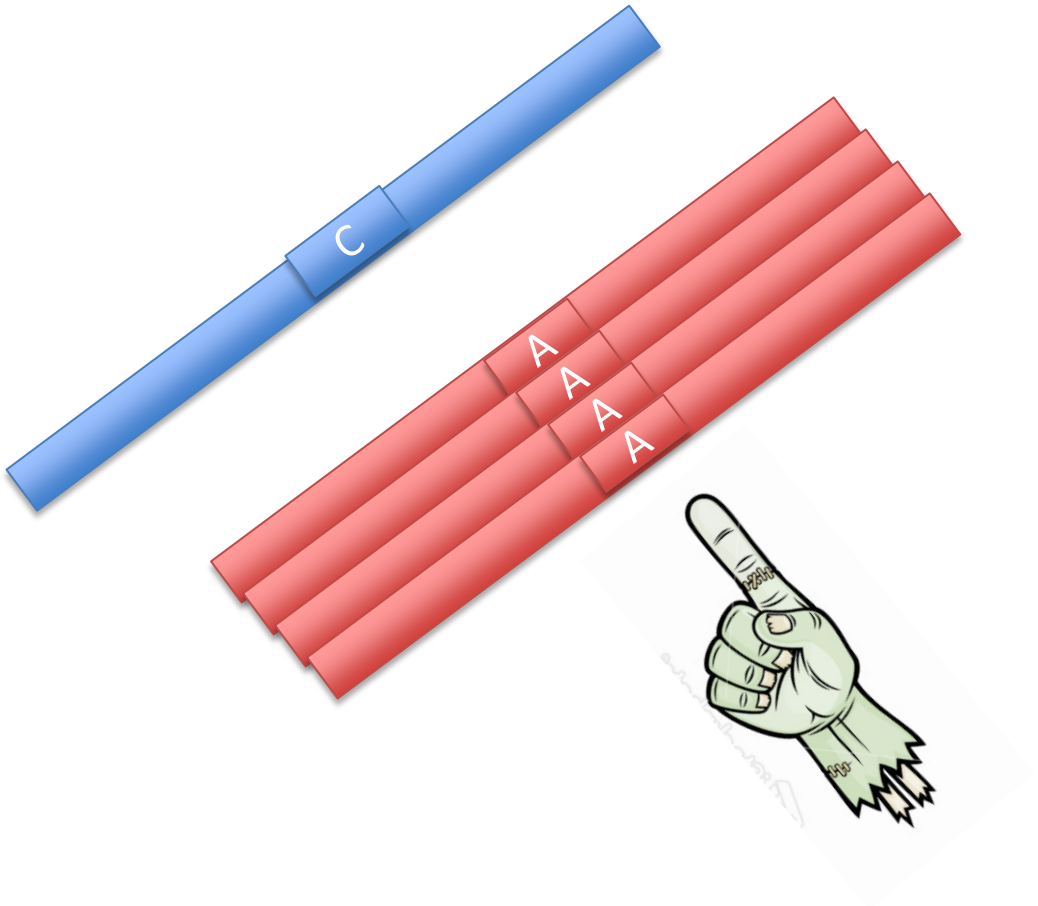
Outi Kilpivaara[1,12], Semanti Mukherjee[2,3,12], Alison M Schram[1], Martha Wadleigh[4], Ann Mullally[4,5], Benjamin L Ebert[5,6], Adam Bass[4,6], Sachie Marubayashi[1], Adriana Heguy[1], Guillermo Garcia-Manero[7], Hagop Kantarjian[7], Kenneth Offit[8], Richard M Stone[4], D Gary Gilliland[4–6,9,10], Robert J Klein[2] & Ross L Levine[1,11]



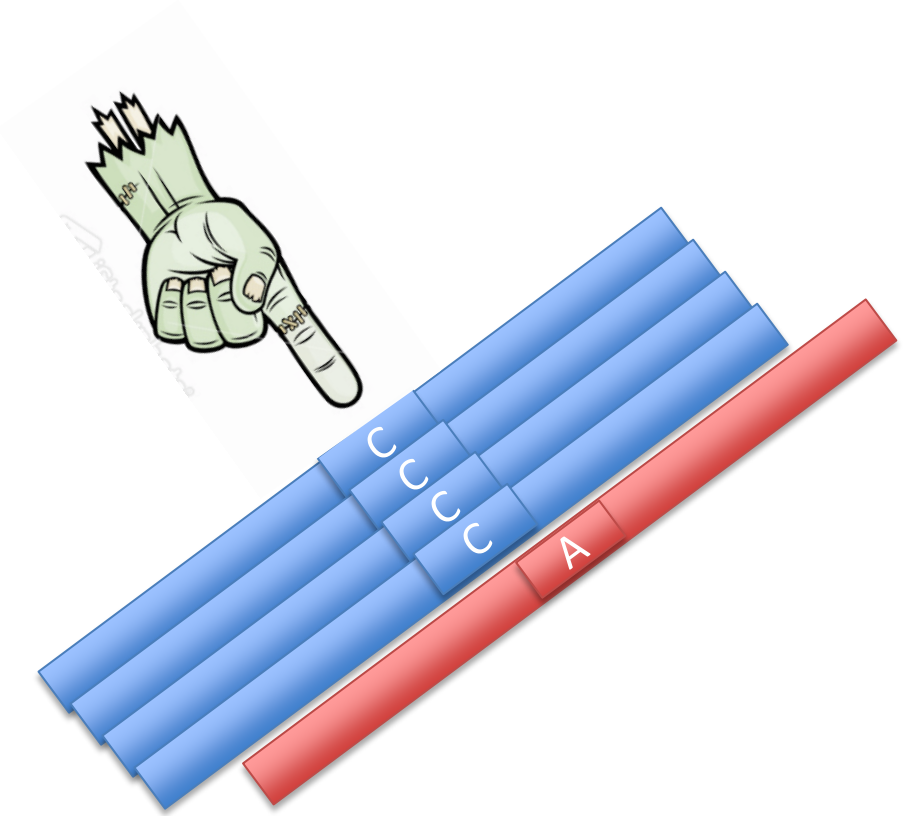
Allelic bias in somatic copy number alterations



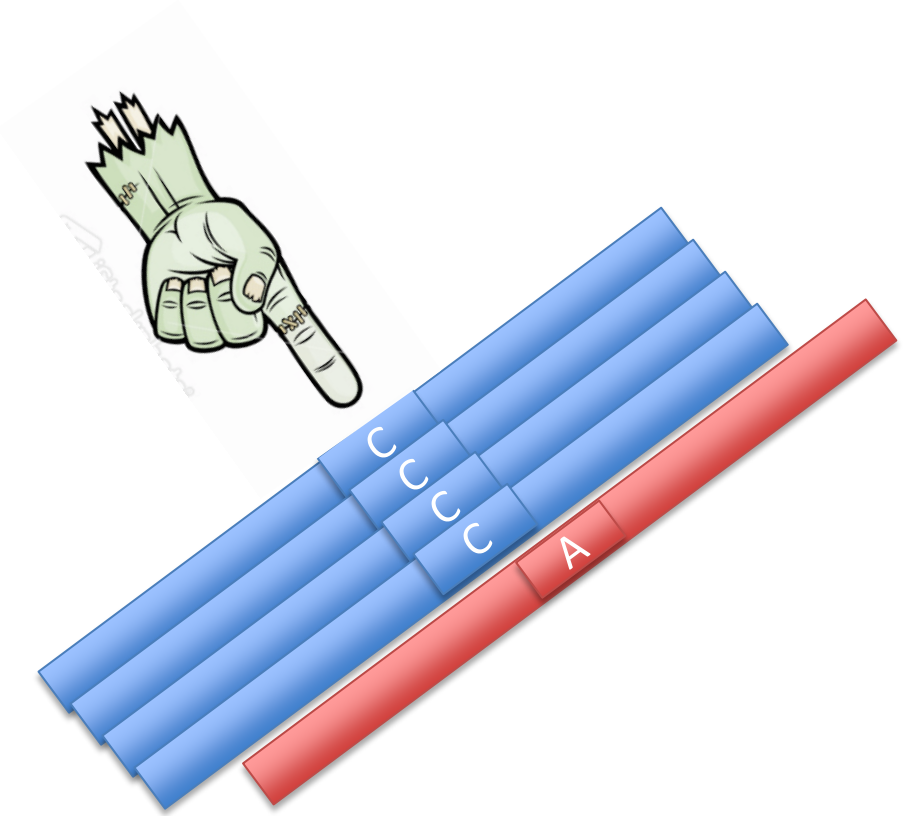
Allelic bias in somatic copy number alterations



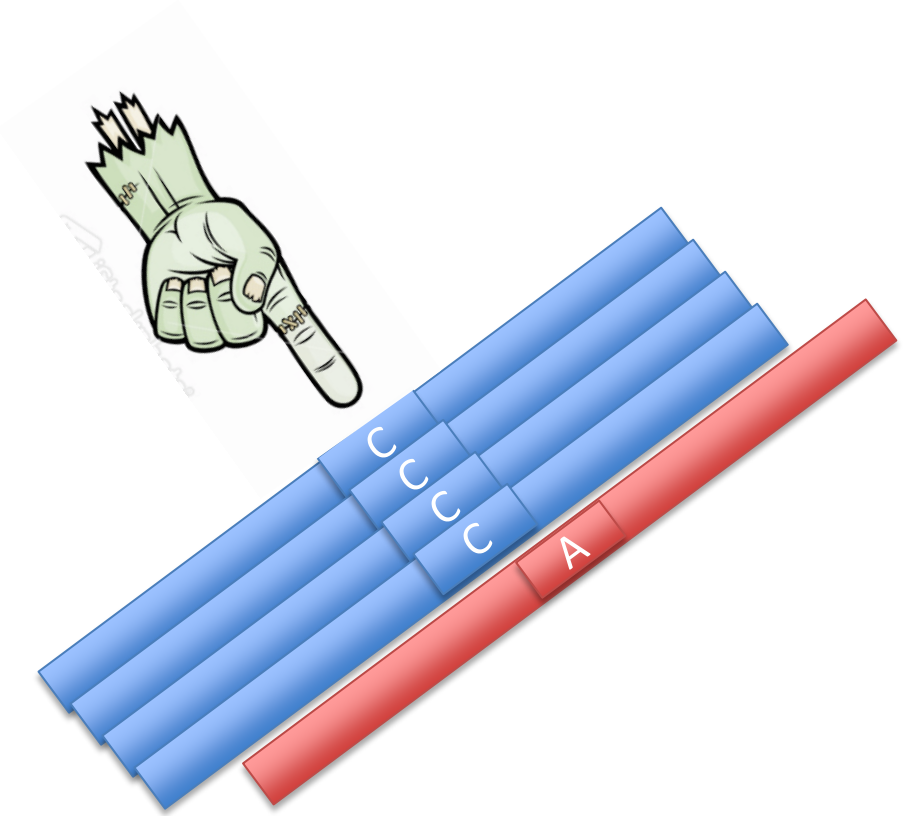
Allelic bias in somatic copy number alterations



Allelic bias in somatic copy number alterations



Allelic bias in somatic copy number alterations



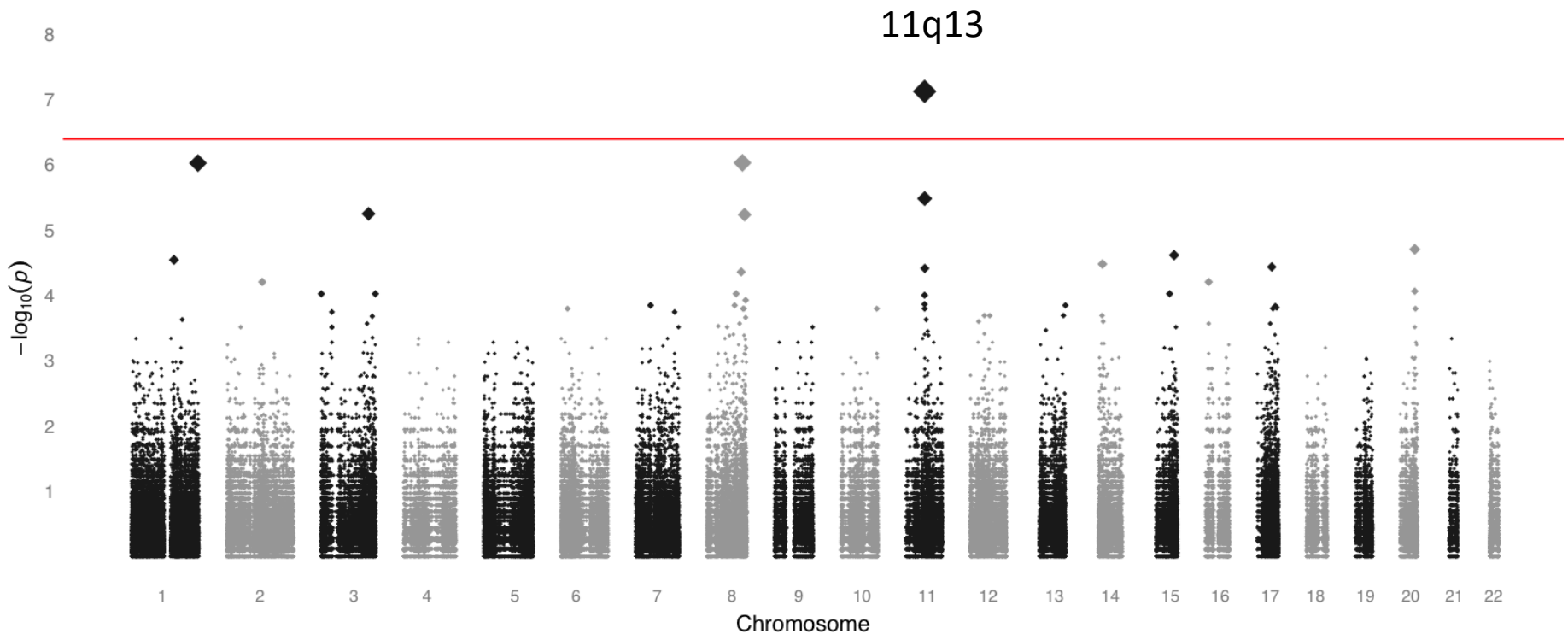
Allelic distortion test (ADT)

- At each heterozygous SNP measure how frequently allele A vs allele B is amplified (or deleted)
- Test significant deviation of frequency from 0.5 via chi square distribution

Allele Amplified	# hets
A	30
a	5

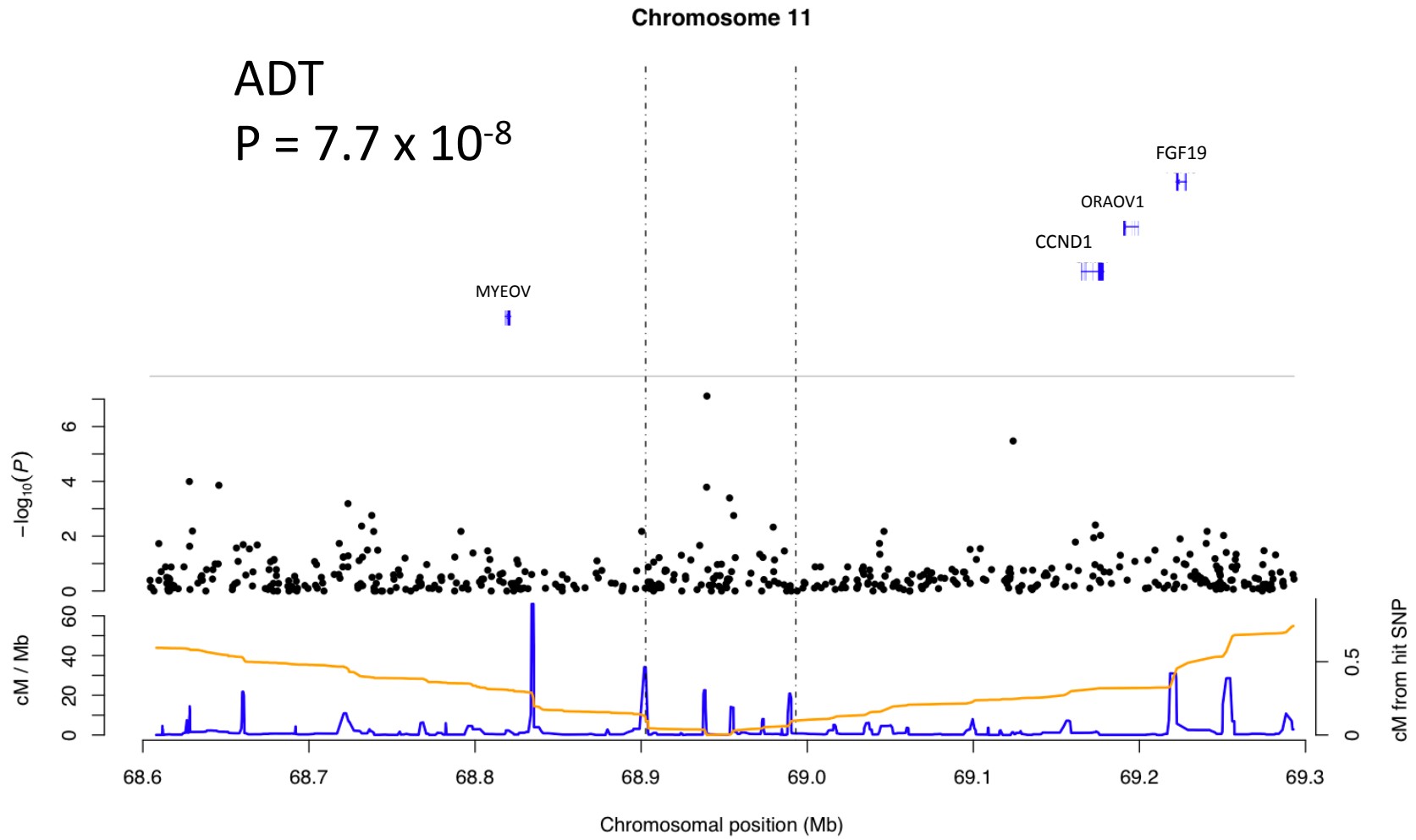
Dewal et al Bioinformatics 2009

Genome-wide ADT
(GCM 250K Affy data, 2643 tumors)

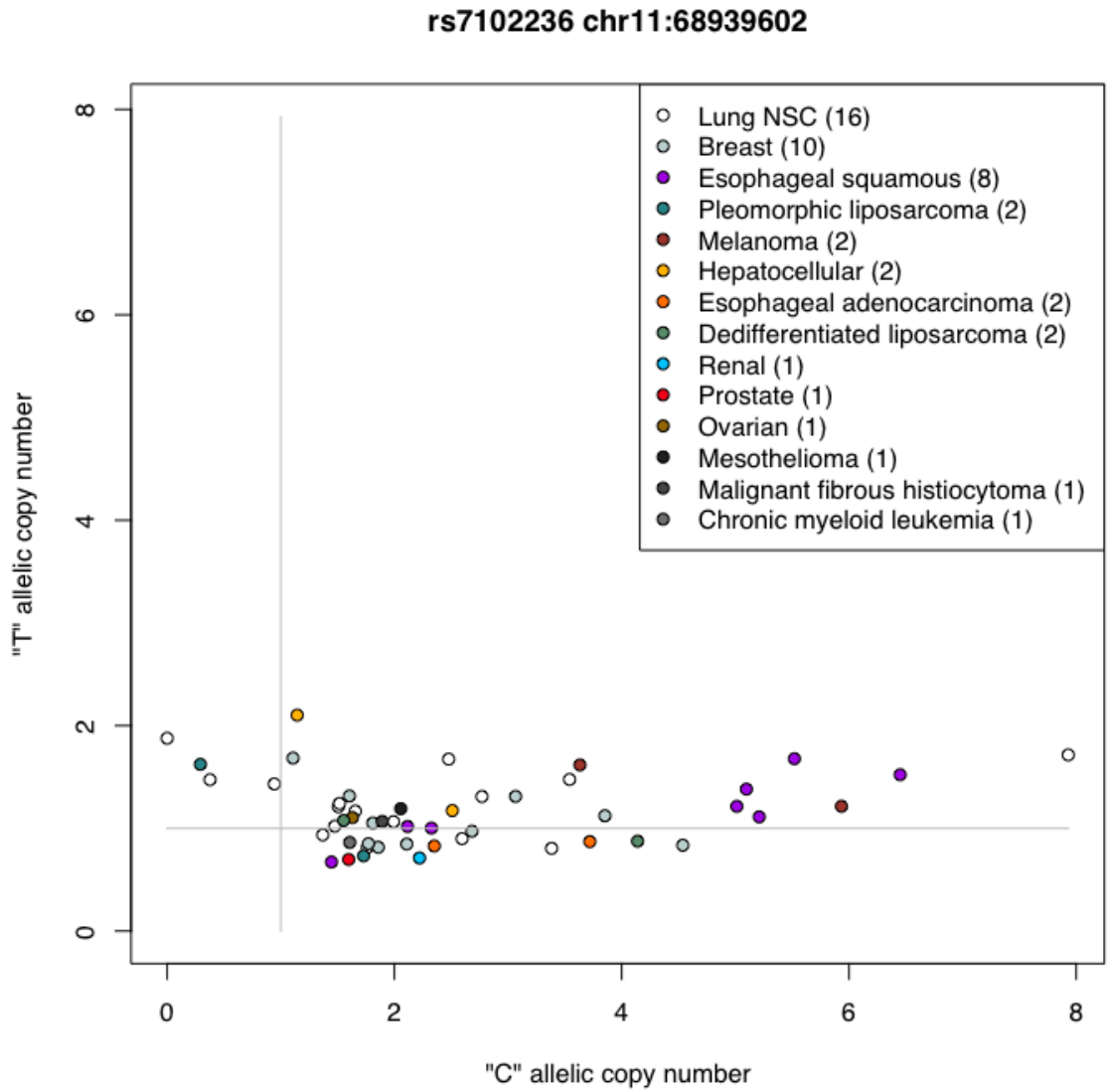


0/36 cancer-GWAS loci intersecting a somatic amplification peak region show significant allelic-distortion

GCM: CCND1 Locus



GCM: CCND1 Locus

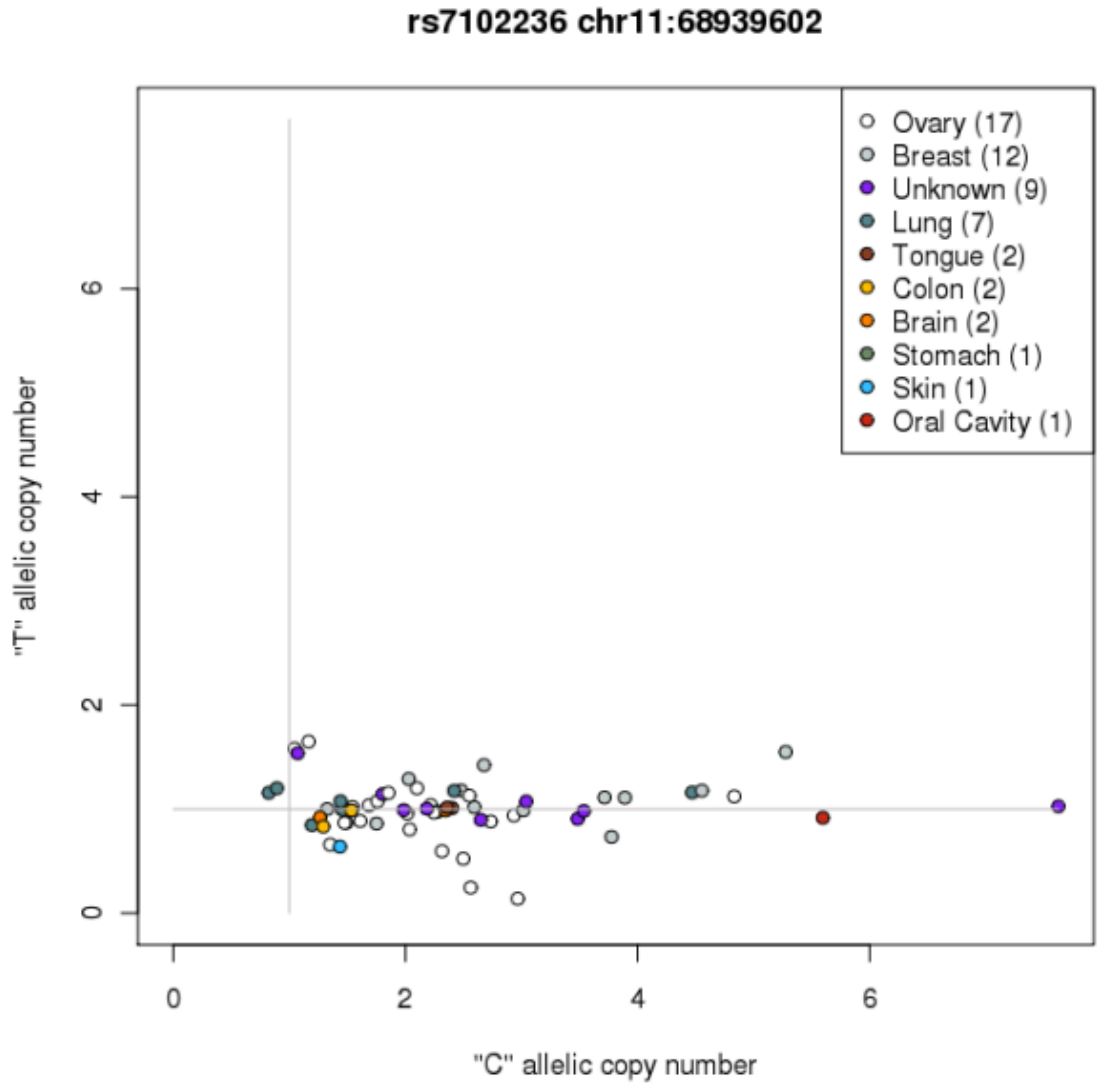


Allele Amplified	# hets
C	44
T	6

ADT

$P = 7.7 \times 10^{-8}$

TCGA 6.0 SNP data: CCND1 Locus

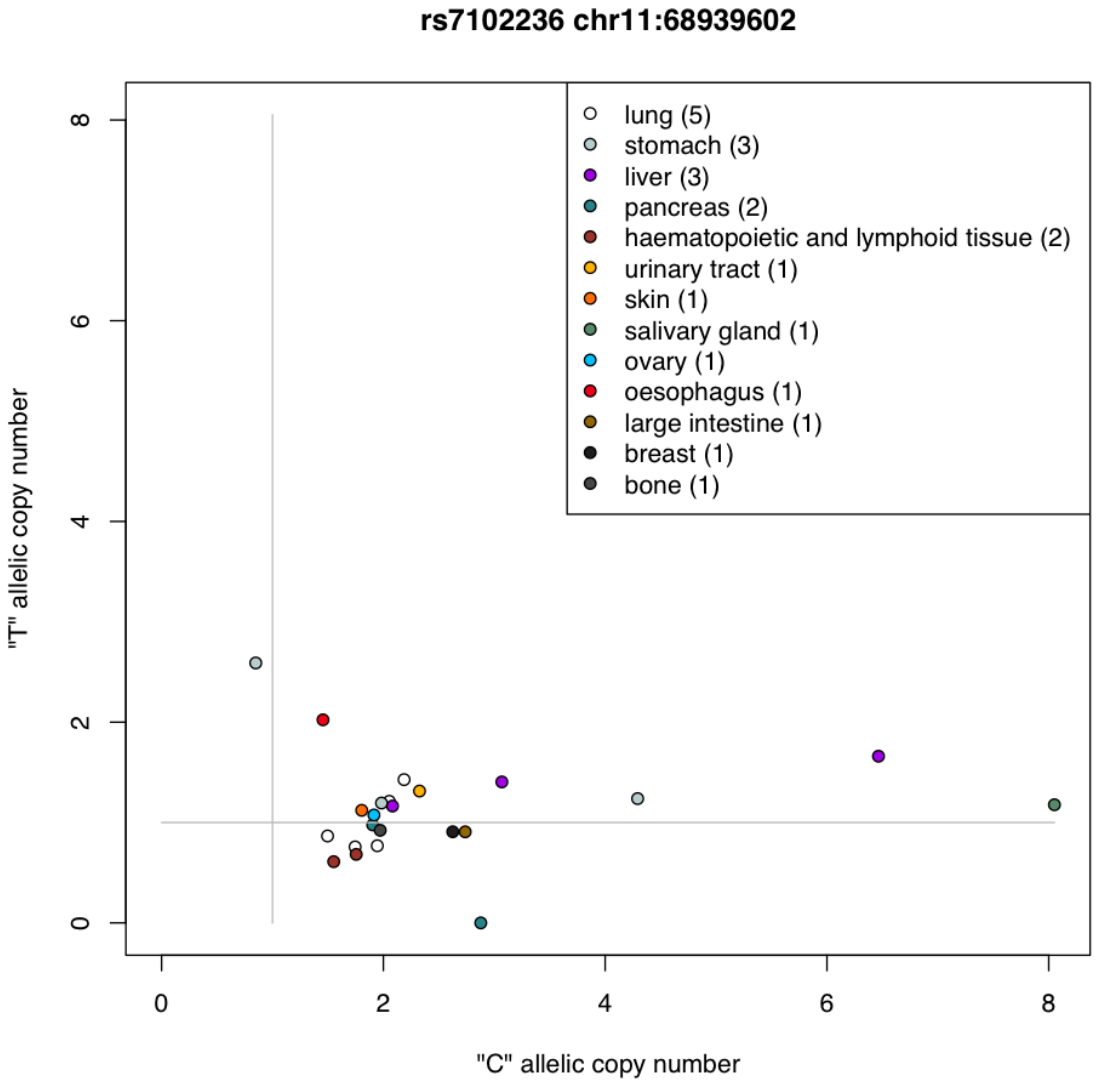


Allele Amplified	# hets
C	57
T	4

ADT

$P = 1.2 \times 10^{-11}$

Broad-Novartis CCLE: CCND1 Locus



Allele Amplified	# hets
C	21
T	2

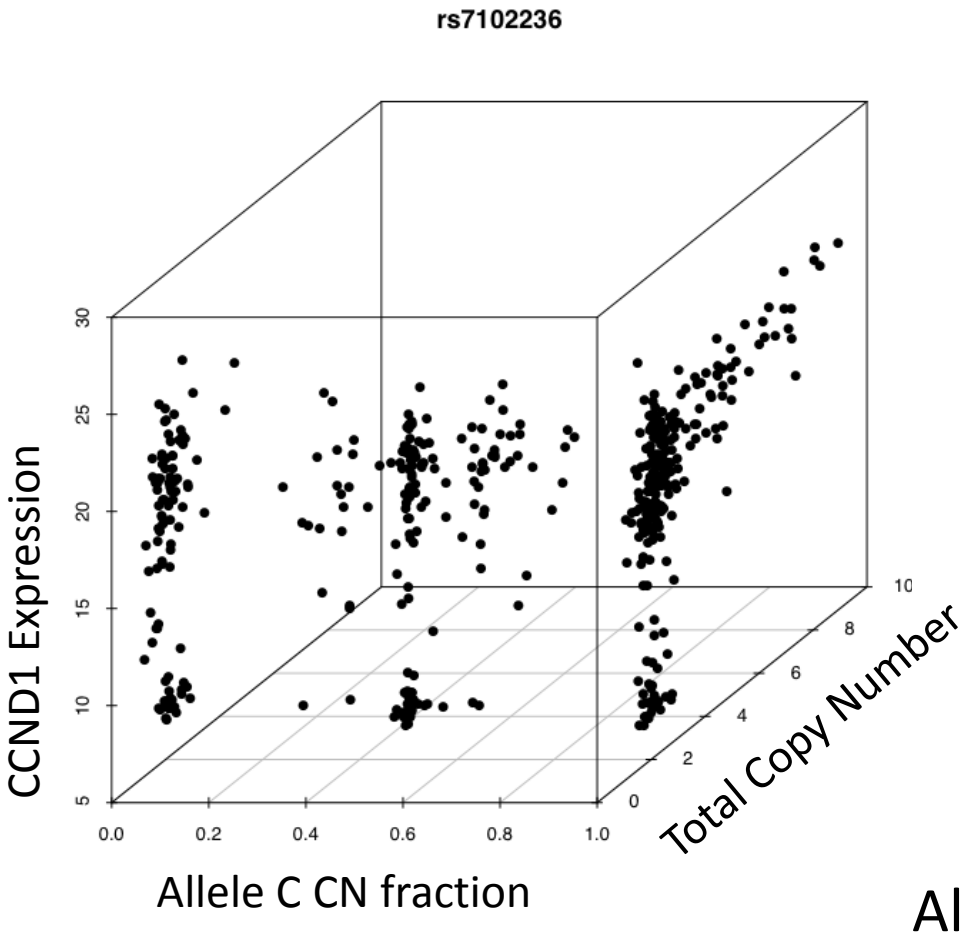
ADT

$P = 7.4 \times 10^{-5}$

Biological Significance?

- Tumors prefer to amplify the C allele at rs7102236, 100KB upstream of CCND1. Why?
- Does the C allele carry a selective advantage?
  - e.g. carry a more “expressible” cyclin D1? or a more “selectively advantageous” cyclin D1?
- Does the C allele just “like to be amplified” more ..
  - ?? interacting with some “amplification” machinery in some “background” way

Broad-Novartis CCLE: CCND1 eQTL



Model:

CCND1 Expression =
Allele C Fraction
+ Total CN

Allele C Effect:
P = 0.0045

Summary

- We see significant overlap of germline GWAS peaks and SCNAs (amps, amps + dels) across cancer types
- First evidence for genome wide colocalization of germline susceptibility variants and somatically altered loci
- CCND1 SNP is a candidate cis-STL (“somatic trait locus”)

Acknowledgements

Scott Carter (Broad)

Rameen Beroukhim (Broad, DFCI)

Craig Mermel (HMS)

Gaddy Getz (Broad)

Matthew Meyerson (Broad, DFCI)