

ENCODE & Cancer

Cancer AWG Sub-working group

3.14.15 (Pi Day)

Mark Gerstein

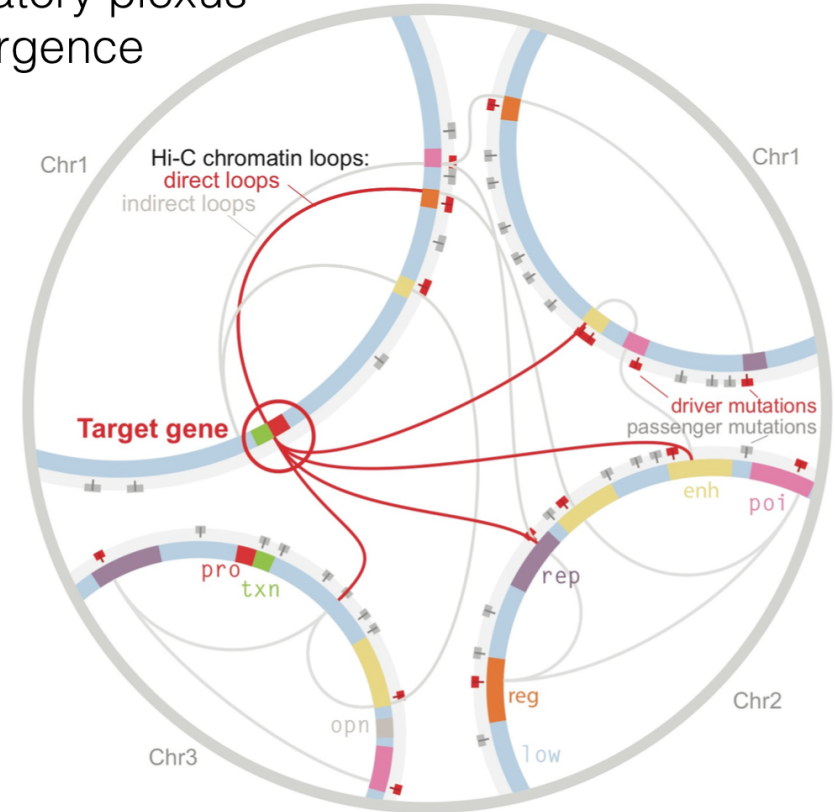
Cancer Group activities:

- 1) **Outreach:** Making ENCODE annotation useful to the cancer world.
- 2) **Outreach:** Harmonizing pipelines.
- 3) **Scientific presentations:** Highlighting Specific ENCODE results on cancer variants.
- 4) **Inreach:** learning about aspects of cancer genomics relevant to ENCODE

Outreach: Making ENCODE annotation useful to the cancer world: Plexus (Kellis Lab)

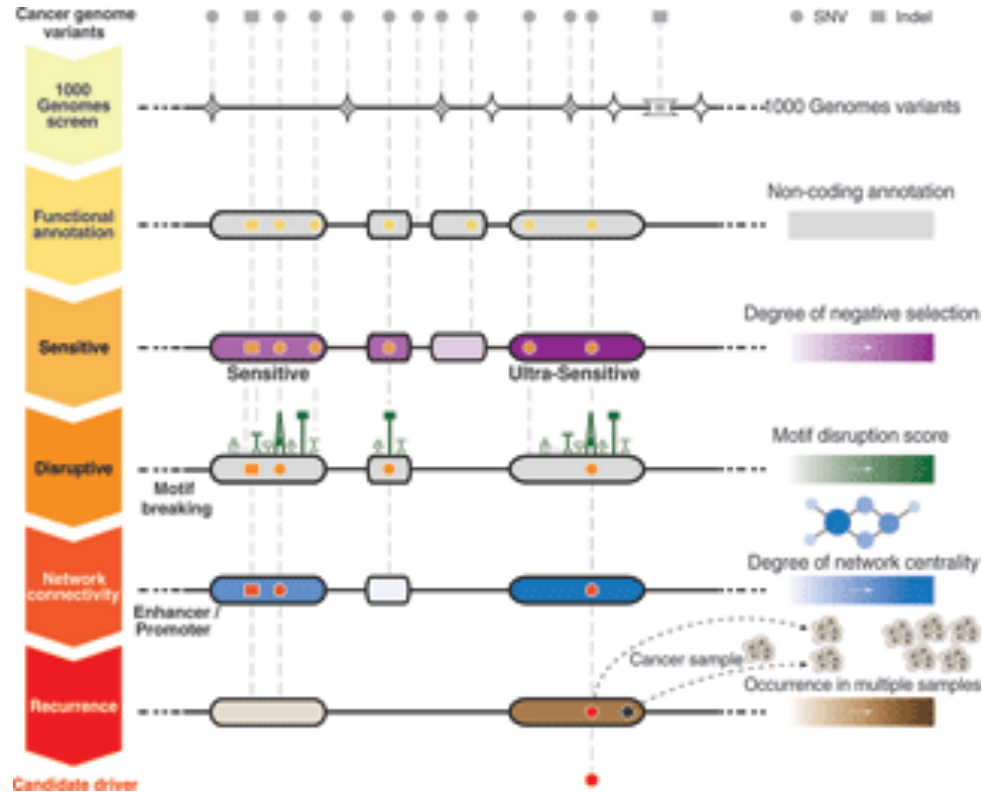
- Plexis are used in a permutation test of the significance of mutational burden.

Regulatory plexus convergence model



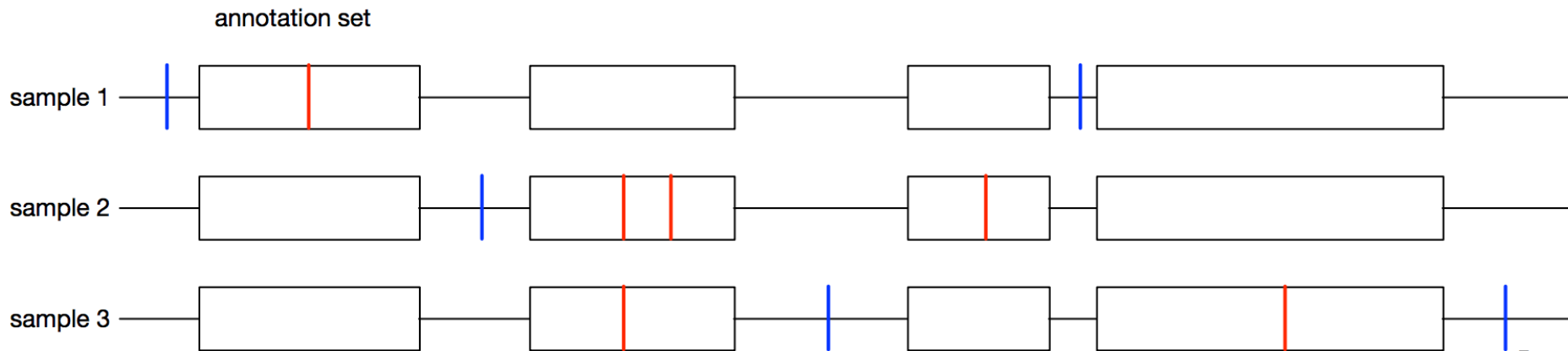
Outreach: Making ENCODE annotation useful to the cancer world: FunSeq (Gerstein Lab)

- started with 677 high-resolution non-coding categories.
- defined ultra sensitive regions with increased with 400x increase in disease assoc.



Outreach: Making ENCODE annotation useful to the cancer world: LARVA (Gerstein Lab)

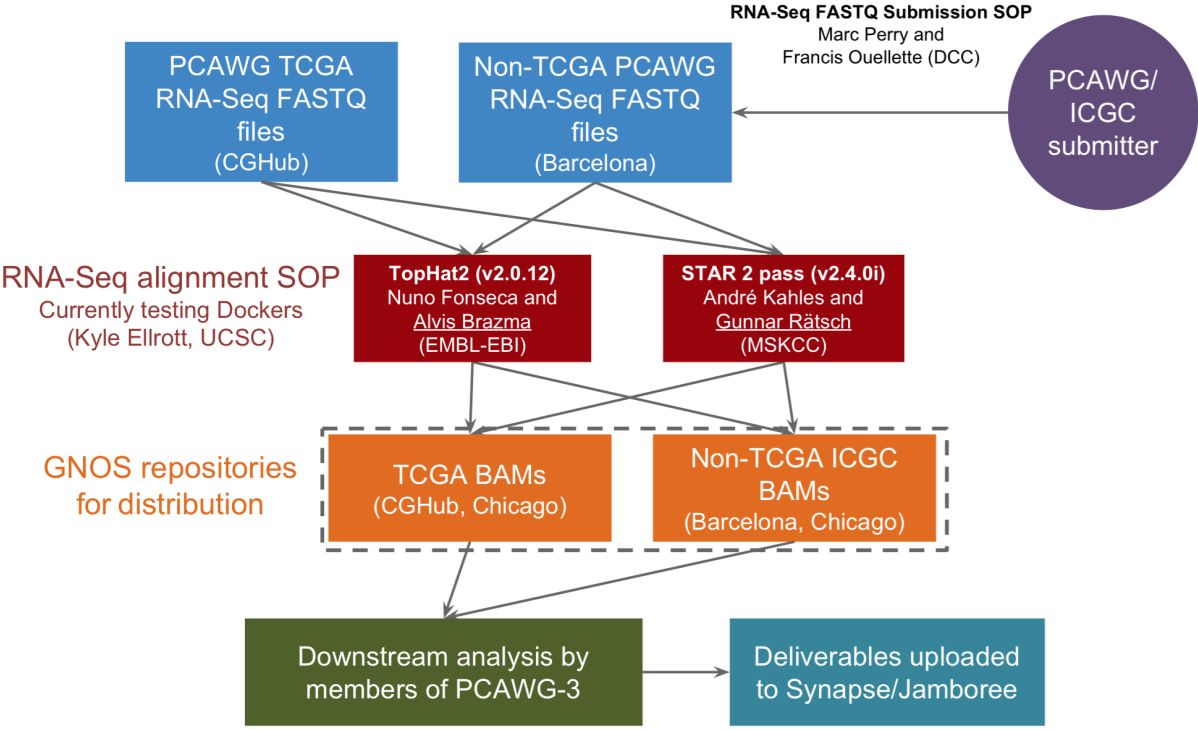
- SNVs from multiple samples were intersected with the ENCODE annotations.
- the significance of the number of SNVs in an element can be then assessed via a permutation test.



Outreach: Harmonizing pipelines: Gunnar Ratsch

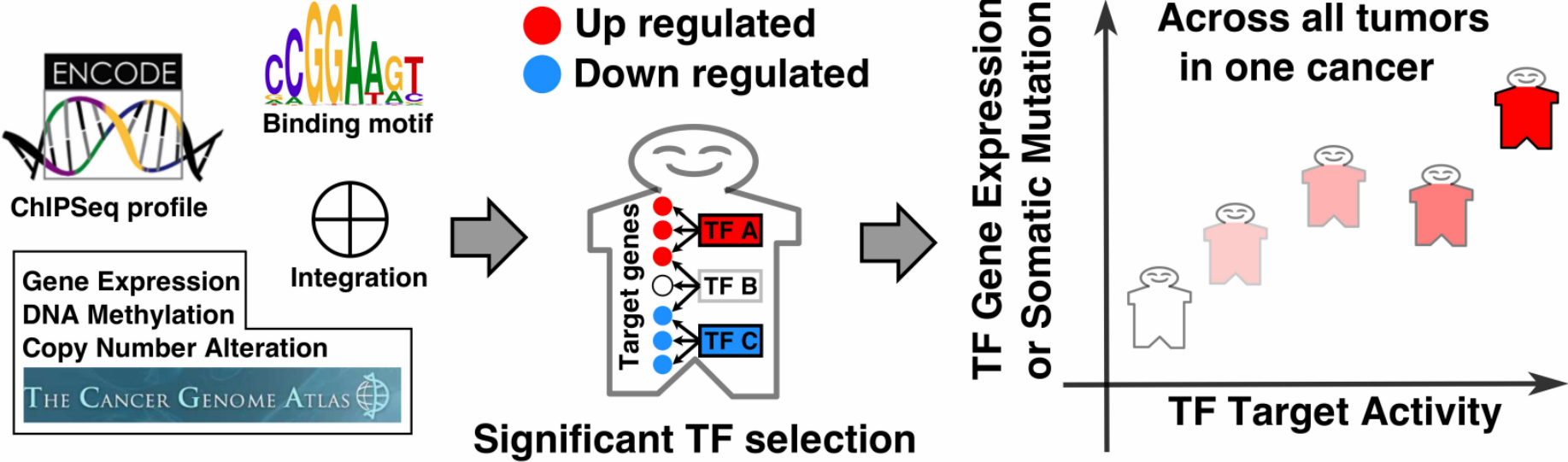
Current Alignment Setup

- Through the use of standard pipelines we can better characterize transcription alteration that cause cancer.



Outreach: Specific results on cancer variants: RABIT (Shirley Liu Lab)

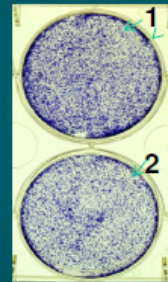
- Pan-cancer assessment of differentially regulated gene expression using penalized regression.



Scientific presentations: Specific results on cancer variants: Leonard Lipovich

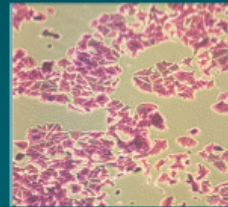
- LncRNAomics has emerged as a clinically relevant field.
- LncRNAs, that are functional in cell growth and cell death, in ER+ breast cancer.

BC041455 is our top estrogen-repressed human lncRNA

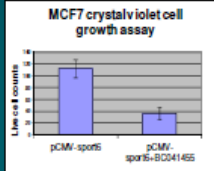
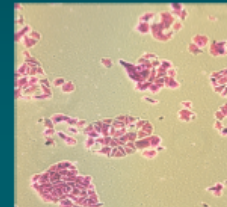


7-17-13 MCF7 cells, seeded 2×10^6 cells/6-well, 120 hours post-OVEREXPRESSION, 96 hours post-E2-treatment.

1 -pCMV-sport6 empty vector



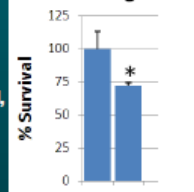
2-pCMV-sport6+BC041455 (OVEREXPRESSION)



MTT assay results:

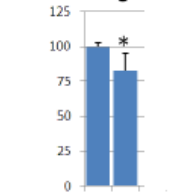
ER α +

MCF7 cells growth



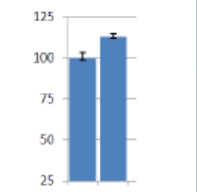
ER α +

T47D cells growth



ER α -

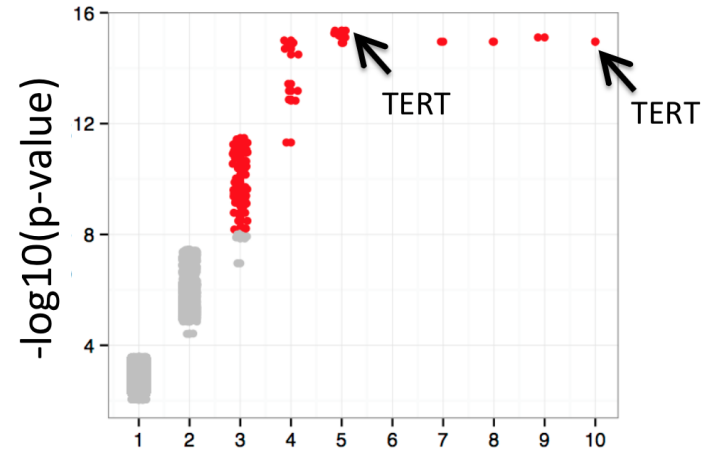
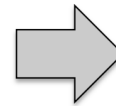
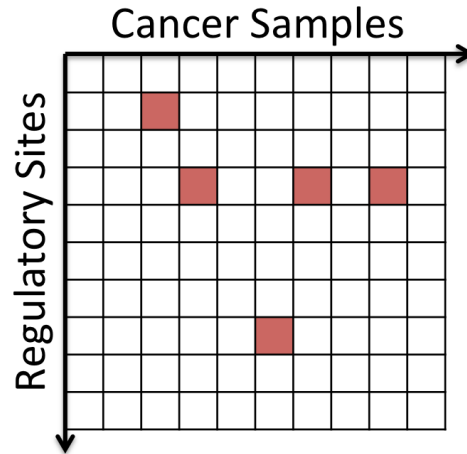
MB231 cells growth



ALL: 120 hours post-OVEREXPRESSION, 96 hours post-E2-treatment.

Scientific presentations: Specific results on cancer variants: TERT (Snyder Lab)

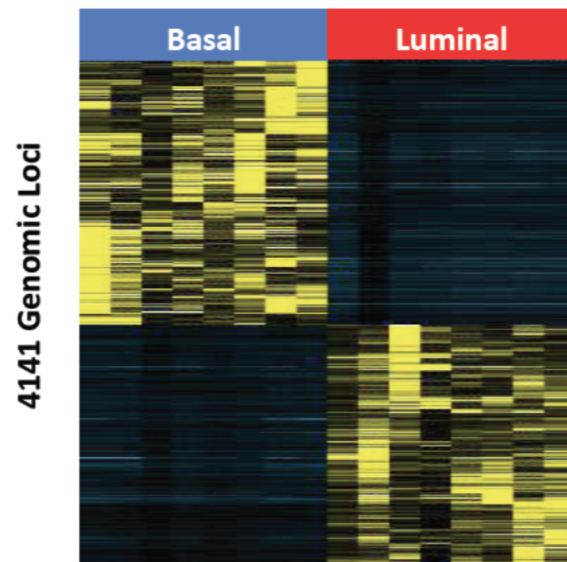
- Observed repeated mutations of positions in regulatory regions selected for in cancer.



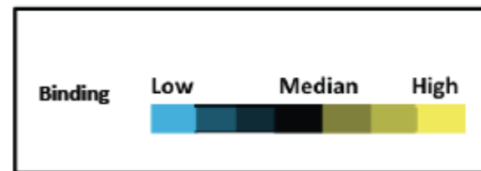
Scientific presentations: Specific results on cancer variants: Breast Cancer (R Myers Lab)

Intersect TF binding sites from ENCODE with genomic regions specifically unmethylated in TNBC (Triple negative breast cancer; basal)

GR and STAT3 binding sites are significantly enriched in unmethylated basal/TNBC loci
Further ChIP-Seq experiments showed that GR and STAT3 bind in a sub-type specific manner

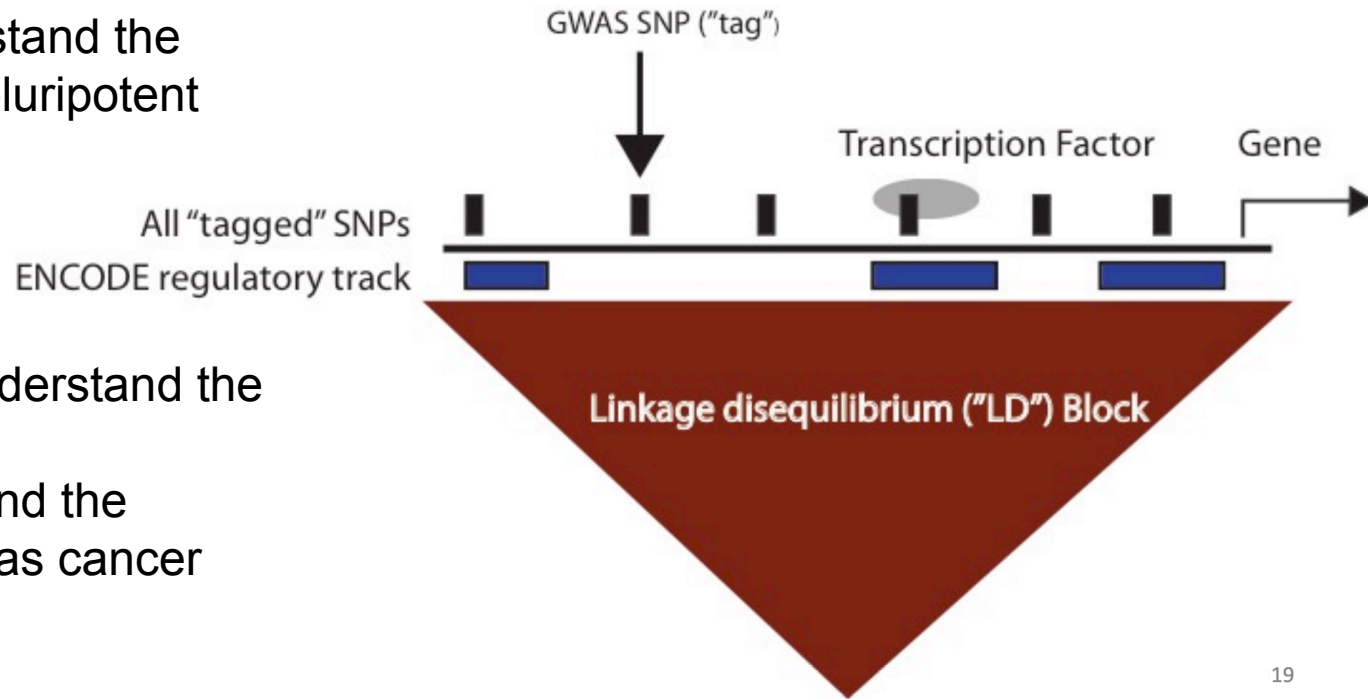


Mann-Whitney-Wilcoxon Test
FDR \leq 0.01



Inreach: Learning about cancer: Robert Klein (ENCODE GWAS)

- Helped us to understand the significance of the pluripotent transformation

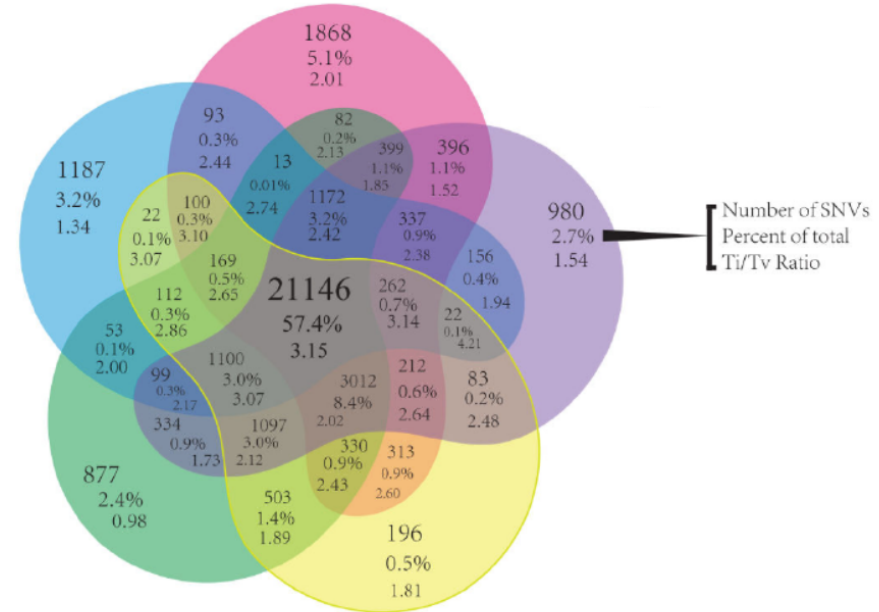


- Helped us better understand the GWAS group.
- Helped us understand the ENCODE cell lines as cancer models

Inreach: Learning about cancer: David Wheeler (from TCGA)

Low concordance in (diploid) SNP calling

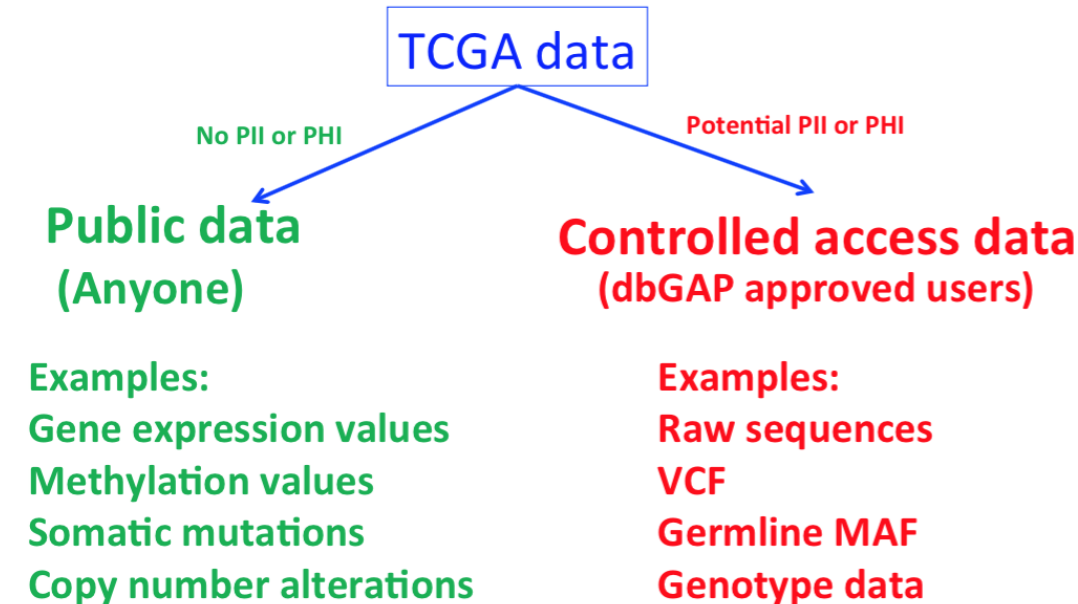
- Helped us to better understand the errors associated with diploid SNP calling in TCGA.



Inreach: Learning about cancer: Zhining Wang (from TCGA program office)

- Helped us better understand what TCGA data that we could use.

Who can get TCGA data?



Specific Outcomes of the Last Year

Identified a group of ENCODE researchers interested in this topic and tools they've put together

Went over access to the data & identified common datasets to focus on (breast/brain, lung/liver)

Discussed a connection with PCAWG either in terms of pipeline harmonization (PCAWG-3) or contributions to analysis

Are we ready to do something more coordinated related to the analysis of cancer data using ENCODE annotation ?