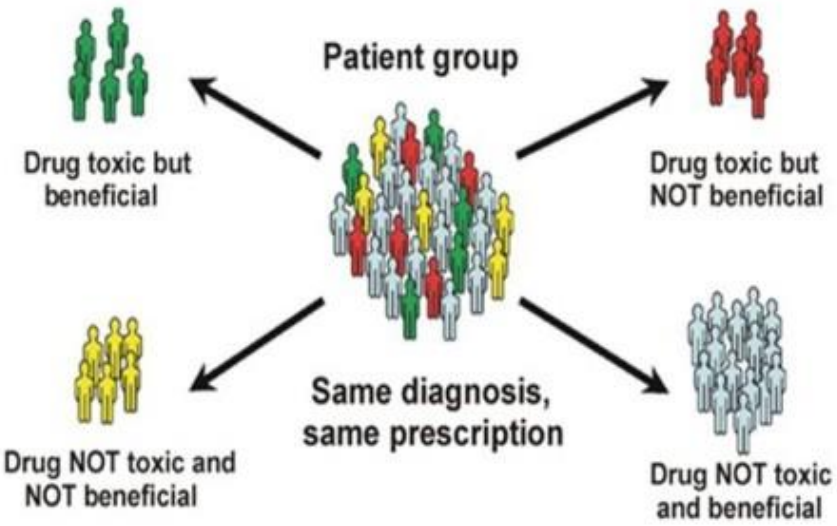
Omics features related to Clinical drug response：a pan-cancer perspective

Zijian Ding

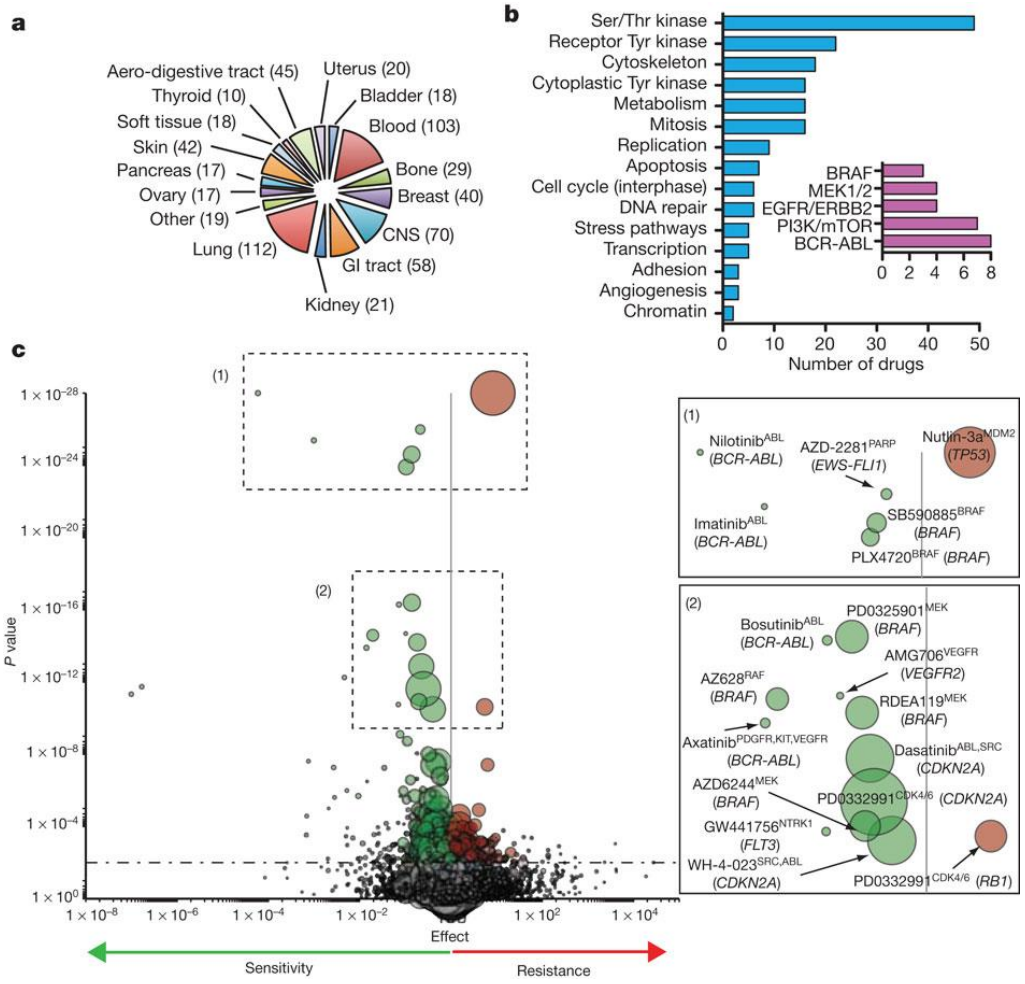
2015.4.3

- Pharmacogenomics (Evans, et.al, *Science*,1999; Evans, et.al, *Nature*, 2003)
  - maximize drug efficacy and minimize toxicity
  - Inherited genetic determinants of drug response
    - Influencing drug disposition
      - Drug metabolism
      - Drug transporters
    - Genetic polymorphisms in drug targets
    - Indirect effects on drug response
  - Leading to “personalized medicine”



Sgugenetics/What is pharmacogenomics

- Cancer genes/ multiple gene signatures are drug-sensitivity biomarkers; (Garnett, et.al, *Nature*, 2012)
- Tumor lineage as predominant marker for some drugs (Barretina et.al, *Nature*, 2012)



(Garnett, et.al, *Nature*, 2012)

Hypothesis: presence of a molecular biomarker predicts response to a targeted therapy independent of tumor histology

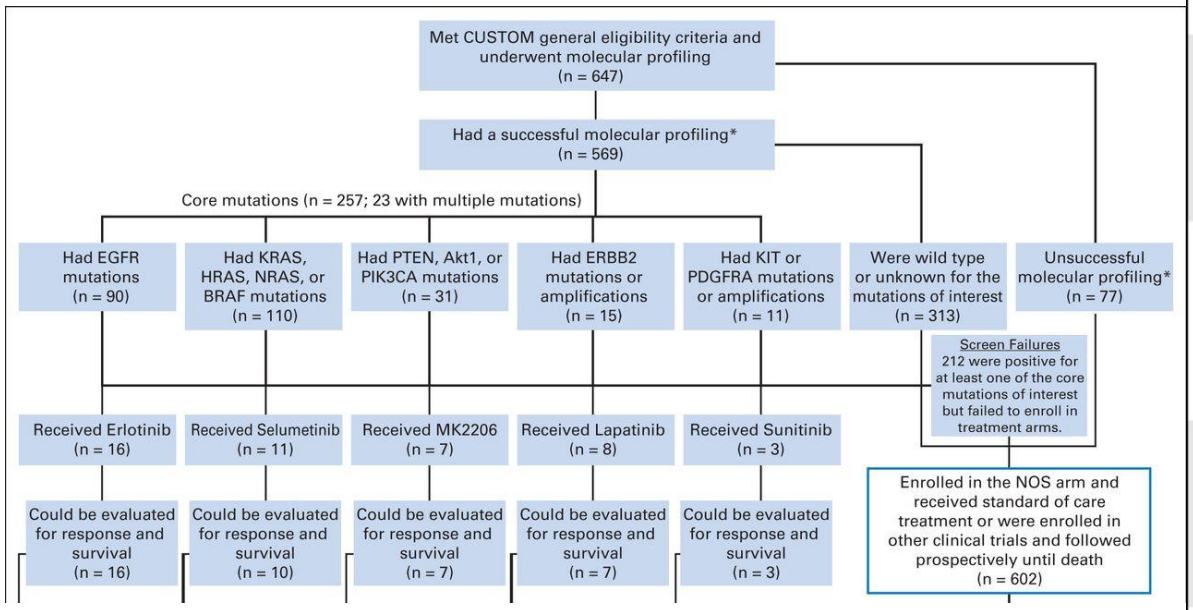


Table 3. Enrollment and Efficacy Assessments

Cancer and Treatment	No. of Patients Enrolled	No. of Patients Evaluable	PR (No.)	SD (No.)	PD (No.)	ORR (%)
NSCLC						
Erlotinib	15	15	9	5	1	60
Lapatinib	7	6	0	4	2	0
Sunitinib	2	2	0	1	1	0
Selumetinib	10	9	1	4	4	11
MK2206	4	4	0	4	0	0
SCLC						
Erlotinib	0	0	0	0	0	0
Lapatinib	1	1	0	1	0	0
Sunitinib	0	0	0	0	0	0
Selumetinib	1	1	0	0	1	0
MK2206	2	2	0	0	2	0
Thymic malignancies						
Erlotinib	1	1	0	0	1	0
Lapatinib	0	0	0	0	0	0
Sunitinib	1	1	0	1	0	0
Selumetinib	0	0	0	0	0	0
MK2206	1	1	0	1	0	0

Abbreviations: NSCLC, non–small-cell lung cancer; ORR, overall response rate; PD, progressive disease; PR, partial response; SCLC, small-cell lung cancer; SD, stable disease.

(Lopez-Chavez et.al, *Journal of Clinical Oncology*, 2015)

- Pan-cancer drug response analysis using multi-omics data
  - Drug sensitivity/response analysis is confined to one particular tumor type (Costello, *Nature Computational biology*, 2014)
    - Of individual data types, mRNA expression by microarray have consistently best predictive power
  - Mutational genes and SCNA genes emerge across cancer types (Kandoth, et.al, *Nature*, 2013; Zack, et.al, *Nature genetics*, 2013)
  - Driver mutational or SCNA genes across tumor types can be evidences of prescription (Rubio-Perez, et.al, *Cancer Cell*, 2015)
- The Cancer Genome Atlas
  - A great opportunity to evaluate commonly used anti-cancer drug across different types of cancer with more samples
  - Multi-omics data provide much more comprehensive molecular information than single platform

- **Pan-cancer drug prescription and response records**
  - Anti-cancer drugs across cancer types
  - Multi-omics data of drug-patients
- Pan-cancer Omics features related to clinical drug response
  - Consistently differential CNVs or methylations among different drug response patient groups across cancer types
  - Different responses of patients between patients with and without specific mutations
- Predictive abilities of pan-cancer features
  - CNVs or methylated probes signatures
  - Mutation gene signatures
  - Simultaneous altered genes as biomarkers (Masica, et.al, *Cancer Research*, 2013)

Initial diagnosis

days

Drug1 response
Drug2 response
Drug3 response
…

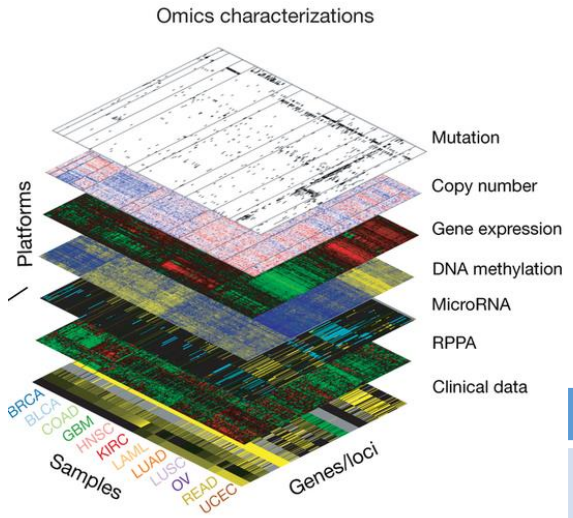
Therapy type: chemo-; Targeted molecular; Hormone; Immuno-

Therapy regimen: Adjuvant; progressive; recurrence

? surgery

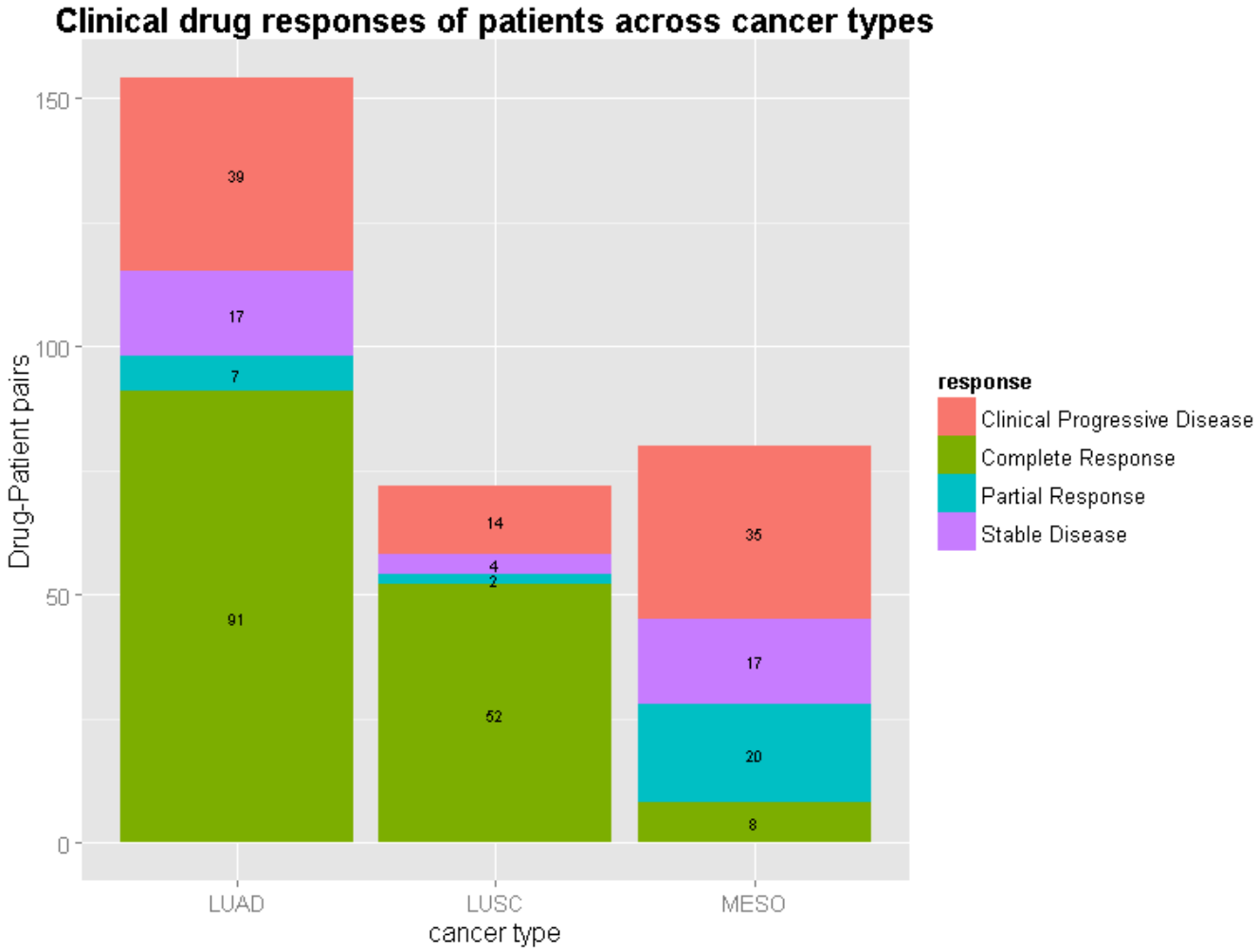
days

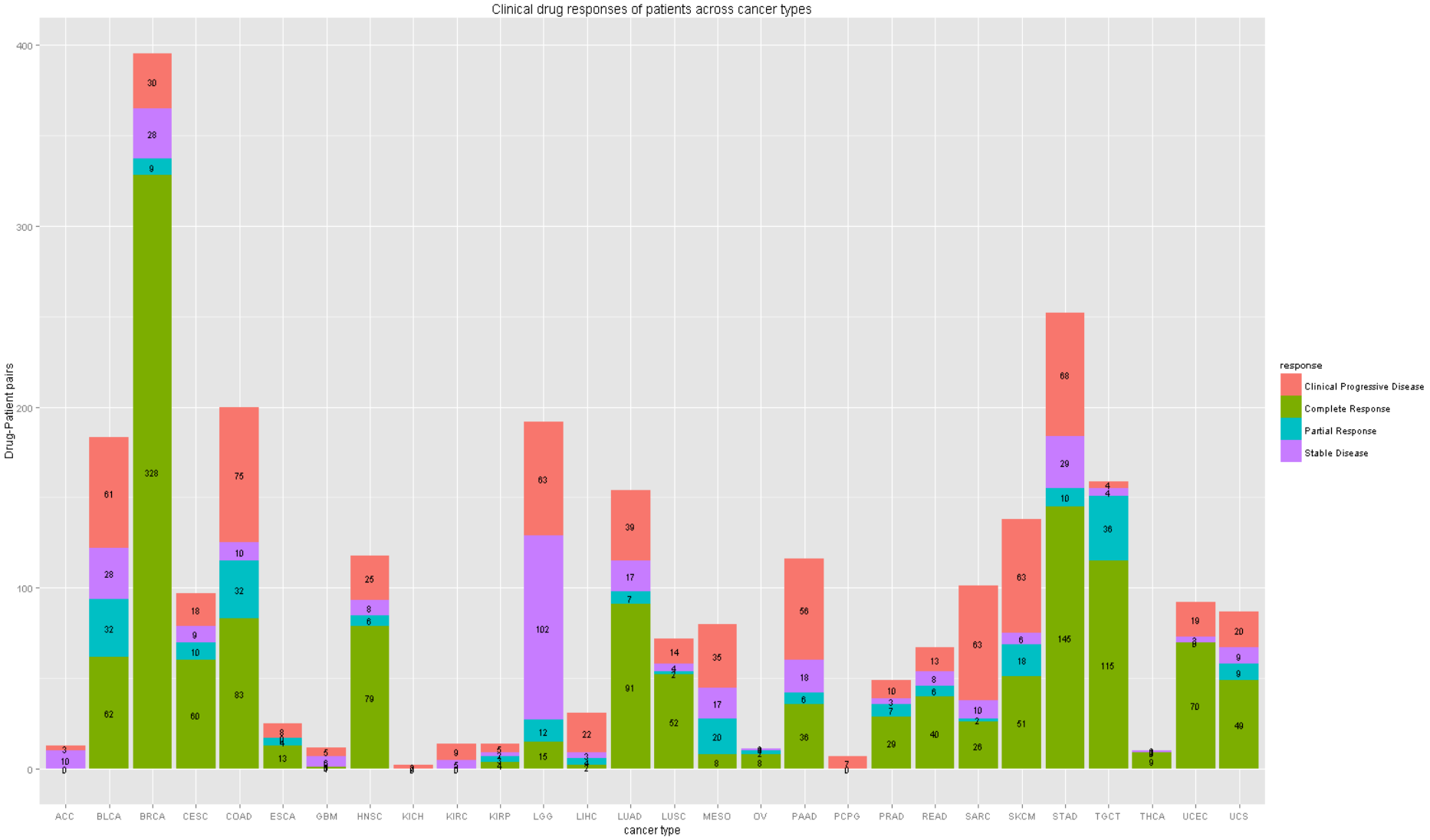
Drug1 response
Drug4 response
Drug5 response

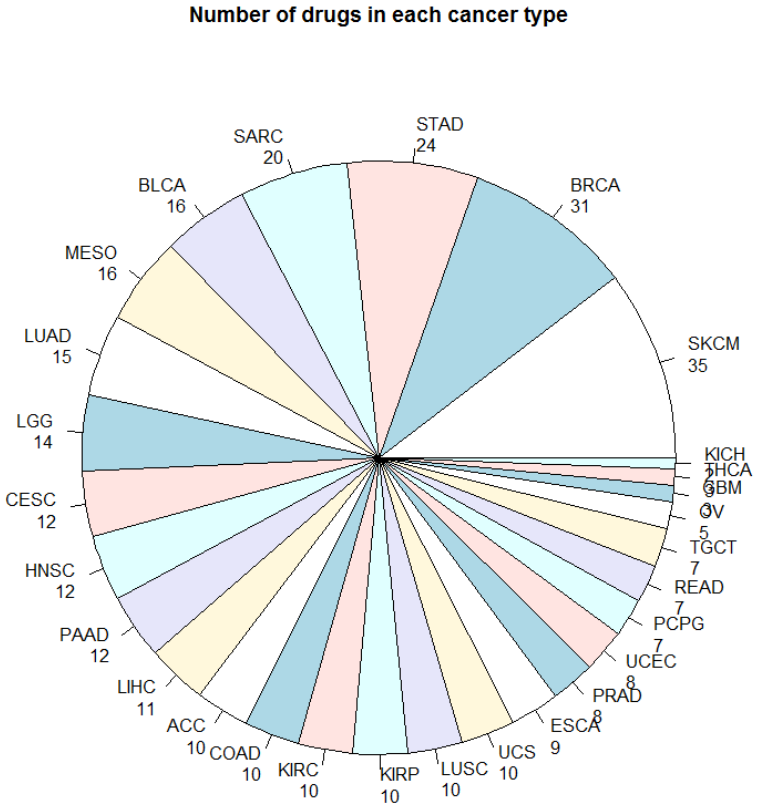


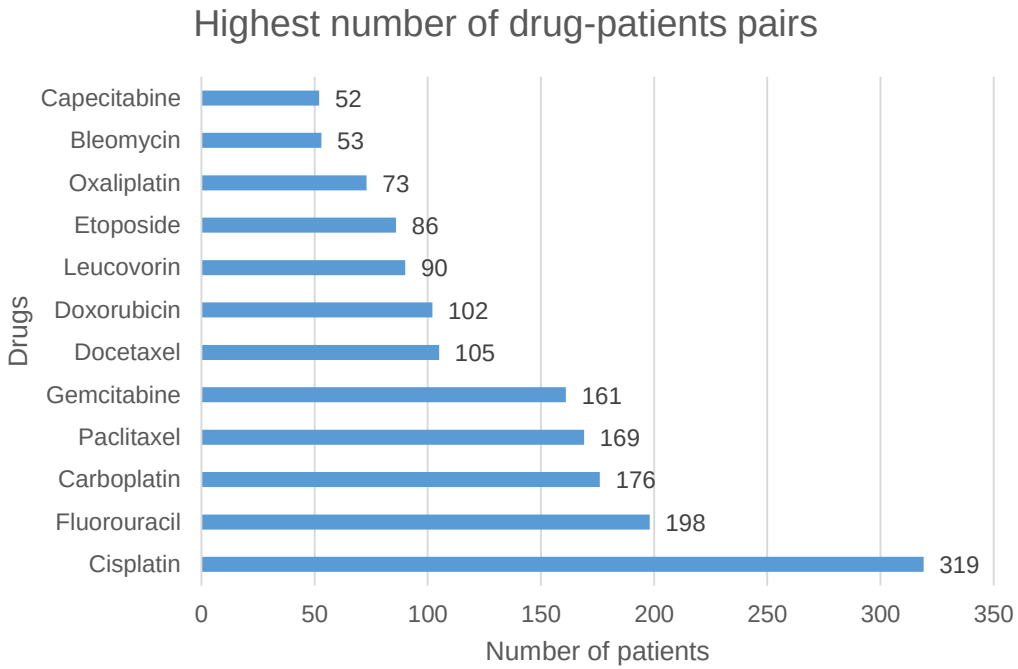
Characterization of drug response
Complete response
Partial response
Stable Disease
Clinical Progressive Disease

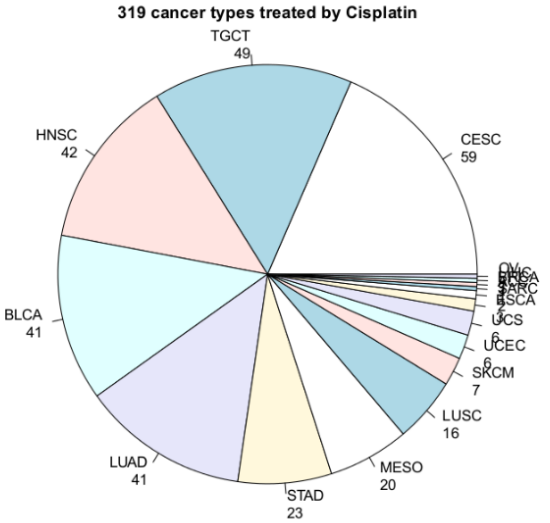
TCGA drug entries	Standard drug names	DrugBank IDs
352	153	99

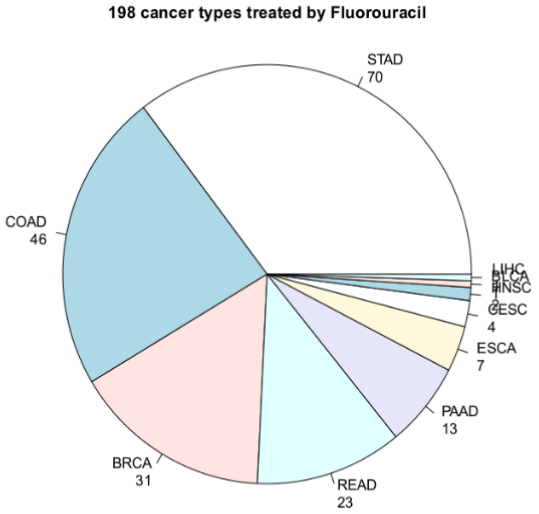

Clinical drug responses of patients across cancer types
Drug-Patient pairs
150
100
50
0
39
17
7
91
14
4
2
52
35
17
20
8
LUAD
LUSC
MESO
cancer type
response
Clinical Progressive Disease
Complete Response
Partial Response
Stable Disease

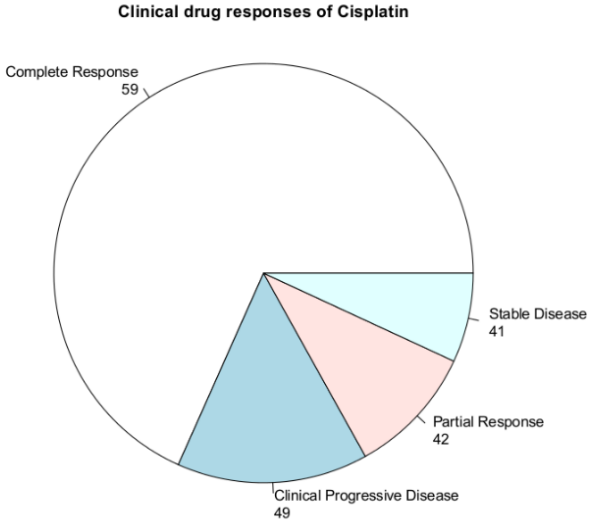

Clinical drug responses of patients across cancer types
Drug-Patient pairs
cancer type
response
Clinical Progressive Disease
Complete Response
Partial Response
Stable Disease
ACC
BLCA
BRCA
CESC
COAD
ESCA
GBM
HNSC
KICH
KIRC
KIRP
LGG
LIHC
LUAD
LUSC
MESO
OV
PAAD
PCPG
PRAD
READ
SARC
SKCM
STAD
TGCT
THCA
UCEC
UCS

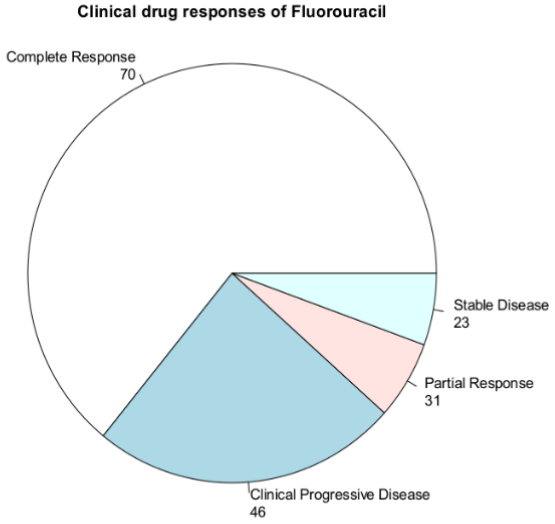

Number of drugs in each cancer type
STAD 24
BRCA 31
SKCM 35
KICH 1
THCA 2
GBM 3
OV 5
TGCT 7
READ 7
PCPG 7
UCEC 8
PRAD 8
ESCA 9
UCS 10
LUSC 10
KIRP 10
KIRC 10
COAD 10
ACC 10
LIHC 11
PAAD 12
HNSC 12
CESC 12
LGG 14
LUAD 15
MESO 16
BLCA 16
SARC 20

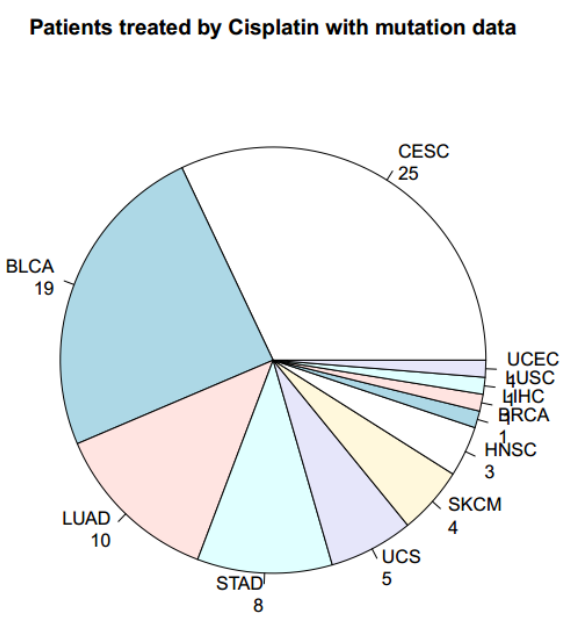

Highest number of drug-patients pairs
Drugs
Capecitabine 52
Bleomycin 53
Oxaliplatin 73
Etoposide 86
Leucovorin 90
Doxorubicin 102
Docetaxel 105
Gemcitabine 161
Paclitaxel 169
Carboplatin 176
Fluorouracil 198
Cisplatin 319
0 50 100 150 200 250 300 350
Number of patients

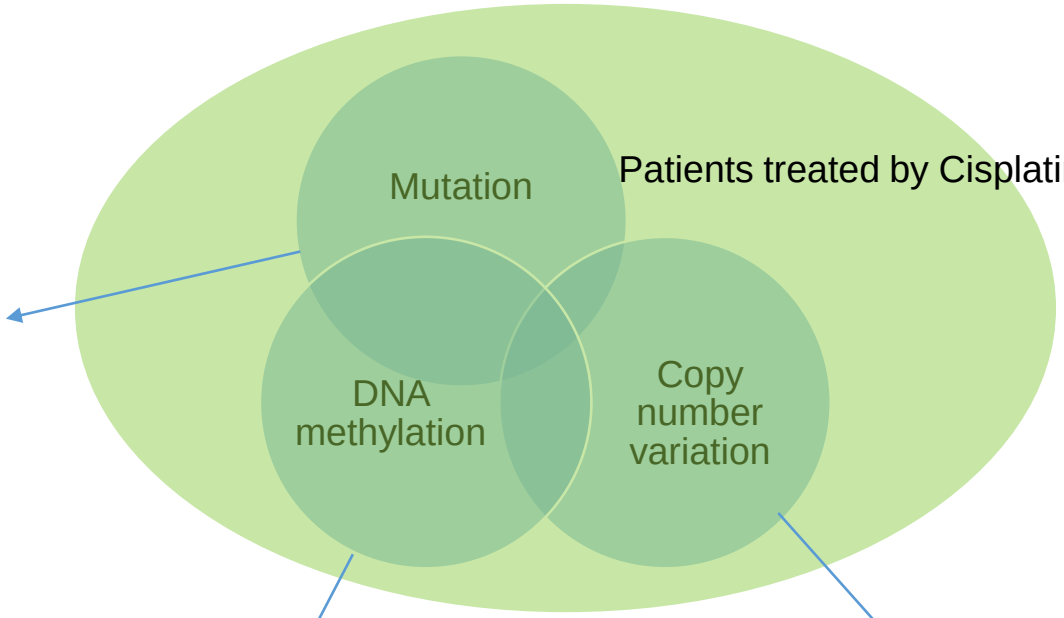

319 cancer types treated by Cisplatin
TGCT
49
CESC
59
HNSC
42
BLCA
41
LUAD
41
STAD
23
MESO
20
LUSC
16
SKCM
7
UCEC
6
UCS
6

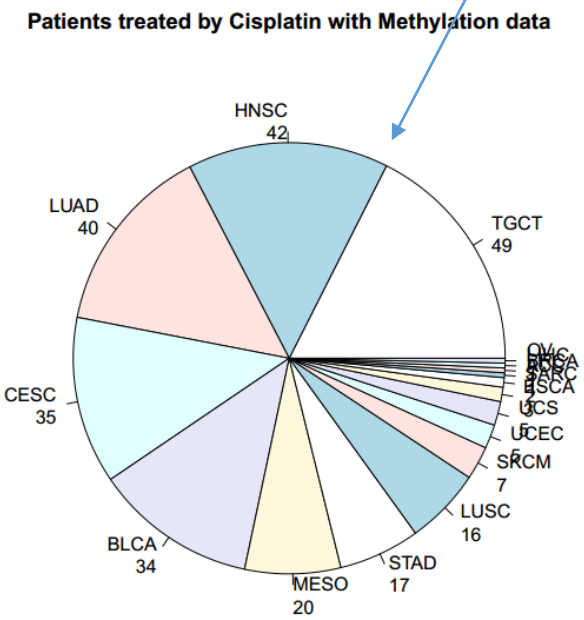

198 cancer types treated by Fluorouracil
STAD
70
COAD
46
BRCA
31
READ
23
PAAD
13
ESCA
7
CESC
4

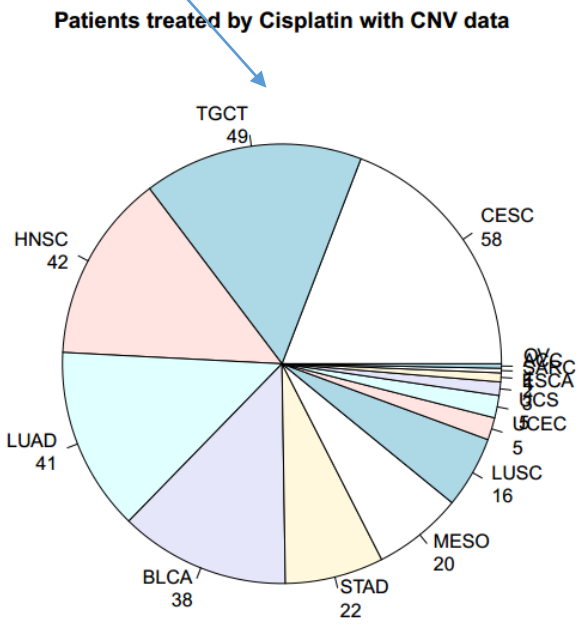

Clinical drug responses of Cisplatin
Complete Response
59
Stable Disease
41
Partial Response
42
Clinical Progressive Disease
49


Clinical drug responses of Fluorouracil
Complete Response
70
Stable Disease
23
Partial Response
31
Clinical Progressive Disease
46


Patients treated by Cisplatin with mutation data
CESC
25
BLCA
19
UCEC
4
UCS
4
LIHC
4
BRCA
1
HNSC
3
SKCM
4
LUAD
10
UCS
5
STAD
8


Patients treated by Cisplatin
Mutation
DNA methylation
Copy number variation

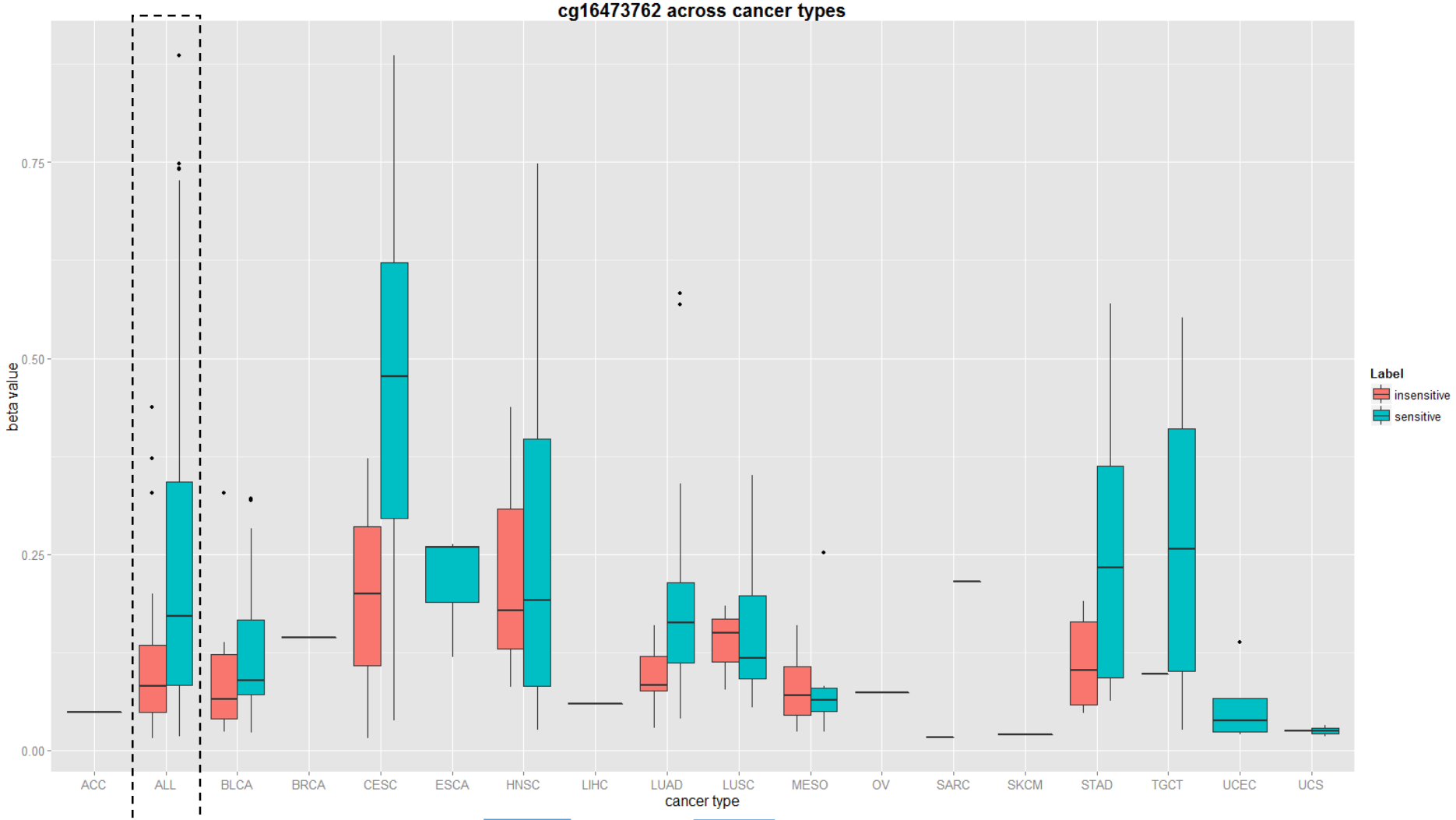

Patients treated by Cisplatin with Methylation data
HNSC
42
LUAD
40
TGCT
49
CESC
35
SKCM
7
LUSC
16
BLCA
34
MESO
20
STAD
17


Patients treated by Cisplatin with CNV data
TGCT
49
HNSC
42
CESC
58
LUAD
41
LUSC
16
MESO
20
BLCA
38
STAD
22

- Pan-cancer drug prescription and response records
  - Anti-cancer drugs across cancer types
  - Multi-omics data of drug-patients

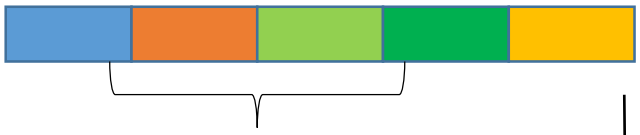
- **Pan-cancer Omics features related to clinical drug response**
  - Consistently differential CNVs or methylations among different drug response patient groups across cancer types
  - Different responses of patients between patients with and without specific mutations

- **Predictive abilities of pan-cancer features**
  - CNVs or methylated probes signatures
  - Mutation gene signatures



diff	0.13	0.04	0.24	0.03	0.09	0.23	0.007	0.14
p	1.4e-13	0.21	0.0012	0.82	0.001	0.6	0.81	0.10

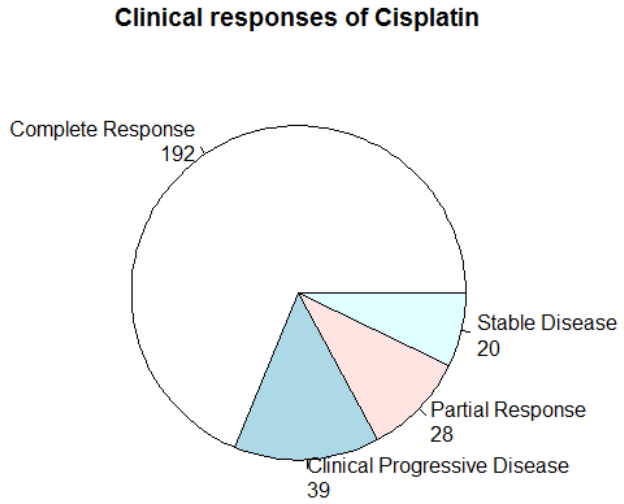
Results Predictive abilities of pan-cancer feature Methylation

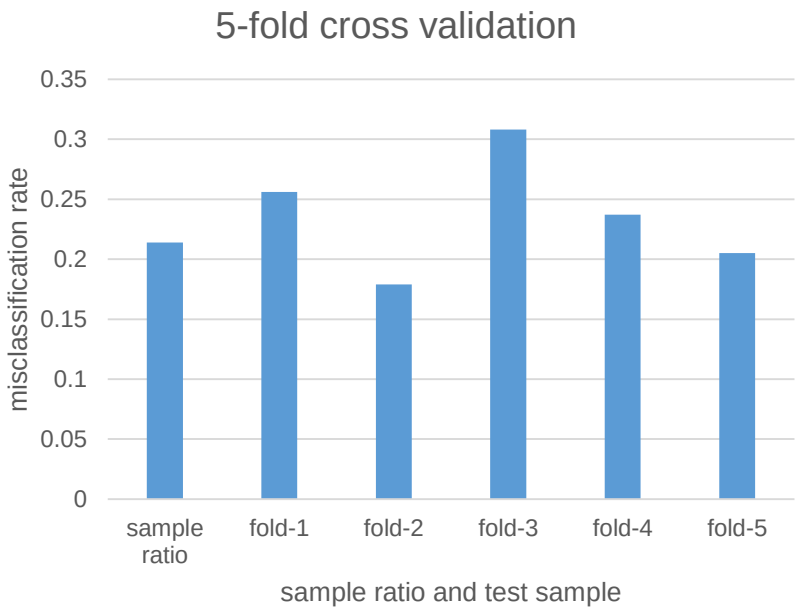


Probe selection: T test

Probe signature selection and model construction: lasso logistic regression

Test predictive power of the probe signature





Discussion

- Pan-cancer Omics features related to clinical drug response
  - Consistently differential CNVs or methylations among different drug response patient groups across cancer types
    - ANOVA: CNV/Methylation ~ response + cancer type
    - Test in each cancer, combine test results
  - Different responses of patients between patients with and without specific mutations
    - ANOVA: response ~ mutation/wild type + cancer type ; (Garnett, et.al, *Nature*, 2012)

Discussion

- Predictive abilities of pan-cancer features
  - Biomarkers/gene signatures selection
    - CNVs or methylated probes signatures
    - Mutation gene signatures
    - Simultaneous altered genes as biomarkers (Masica, et.al, *Cancer Research*, 2013)
  - Predictive model
    - Cancer type as one independent variable

Discussion

- Results validation
  - PharmGKB: SNP/gene-drug response (McDonagh, et.al, *Biomarkers Med*, 2011)
  - GDSC: gene-drug response (Yang, et.al, *Nucleic Acids Research*, 2013)
  - Literature research

A brief introduction to TCGA data

TCGA network

Biospecimen Core Resource (BCR) – Tissue samples are carefully cataloged, processed, checked for quality and stored, complete with important medical information about the patient.

Genome Characterization Centers (GCCs) – Several technologies will be used to analyze genomic changes involved in cancer. The genomic changes that are identified will be further studied by the Genome Sequencing Centers.

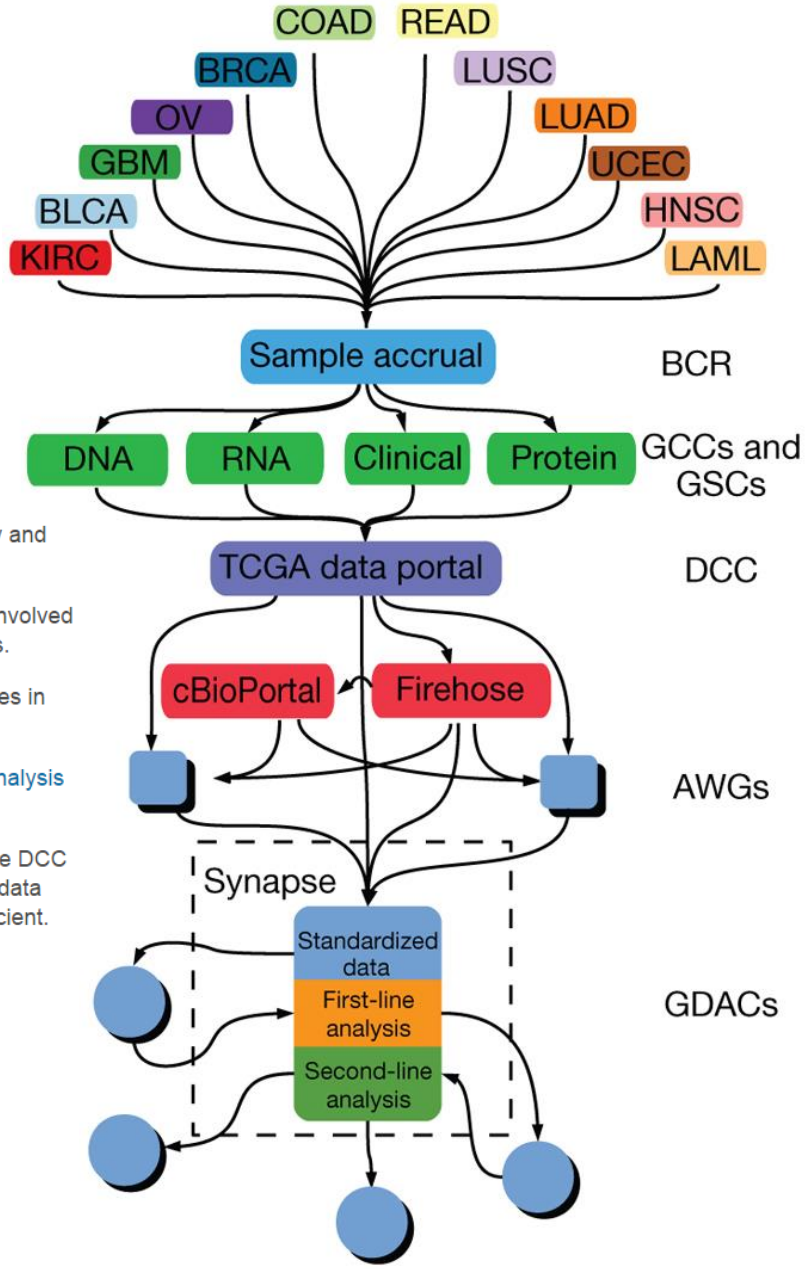
Genome Sequencing Centers (GSCs) – High-throughput Genome Sequencing Centers will identify the changes in DNA sequences that are associated with specific types of cancer.

Proteome Characterization Centers (PCCs) – The centers, a component of NCI's Clinical Proteomic Tumor Analysis Consortium, will ascertain and analyze the total proteomic content of a subset of TCGA samples.

Data Coordinating Center (DCC) – The information that is generated by TCGA will be centrally managed at the DCC and entered into the TCGA Data Portal and Cancer Genomics Hub as it becomes available. Centralization of data facilitates data transfer between the network and the research community, and makes data analysis more efficient. The DCC manages the TCGA Data Portal.

Cancer Genomics Hub (CGHub) – Lower level sequence data will be deposited into a secure repository. This database stores cancer genome sequences and alignments.

Genome Data Analysis Centers (GDACs) – Immense amounts of data from array and second-generation sequencing technologies must be integrated across thousands of samples. These centers will provide novel informatics tools to the entire research community to facilitate broader use of TCGA data.



TCGA network, *Nature Genetics*, 2013

Get Data

```
[zding@console ~]$ cd /data/database/TCGA/Broad/
[zding@console Broad]$ ll
total 8
drwxr-xr-x  3 root  root       31 Feb  8 13:14 Clinical
drwxr-xr-x 34 root  root     4096 Feb 28 15:34 level-3
drwxr-xr-x 35 zding graduate 4096 Mar  3 20:15 level-4
```

```
[zding@console Broad]$ ll level-3/
total 20
drwxr-xr-x 3 root root    31 Feb  8 13:14 ACC
drwxr-xr-x 3 root root    31 Feb  8 13:14 BLCA
drwxr-xr-x 3 root root    31 Feb  8 13:15 BRCA
drwxr-xr-x 3 root root    31 Feb  8 13:17 CESC
drwxr-xr-x 3 root root    31 Feb  8 13:17 CHOL
drwxr-xr-x 3 root root    31 Feb  8 13:17 COAD
drwxr-xr-x 3 root root    31 Feb  8 13:18 COADREAD
-rw-r--r-- 1 root root    97 Feb  7 12:09 data_info.txt
drwxr-xr-x 3 root root    31 Feb  8 13:18 ESCA
drwxr-xr-x 3 root root    31 Feb  8 13:19 GBM
-rw-r--r-- 1 root root   153 Feb 10 09:41 head.qsub
drwxr-xr-x 3 root root    31 Feb  8 13:19 HNSC
drwxr-xr-x 3 root root    31 Feb  8 13:19 KICH
drwxr-xr-x 3 root root    31 Feb  8 13:19 KIRC
drwxr-xr-x 3 root root    31 Feb  8 13:19 KIRP
drwxr-xr-x 3 root root    31 Feb  8 13:19 LAML
drwxr-xr-x 3 root root    31 Feb  8 13:19 LGG
drwxr-xr-x 3 root root    31 Feb  8 13:21 LIHC
drwxr-xr-x 3 root root    31 Feb  9 16:12 LUAD
drwxr-xr-x 3 root root    31 Feb  9 16:12 LUSC
drwxr-xr-x 3 root root    31 Feb  8 13:23 MESO
drwxr-xr-x 3 root root    31 Feb  8 13:23 OV
drwxr-xr-x 3 root root    31 Feb  8 13:24 PAAD
drwxr-xr-x 3 root root    31 Feb  8 13:24 PCPG
drwxr-xr-x 3 root root    31 Feb  8 13:24 PRAD
drwxr-xr-x 3 root root    31 Feb  8 13:26 READ
drwxr-xr-x 3 root root    31 Feb  8 13:26 SARC
```

```
[zding@console Broad]$ ll level-3/BLCA/2014.12.06/ -t
total 2530000
-rw-r--r-- 1 root root      32044 Feb 28 16:58 head.qsub.o356631
drwxr-xr-x 2 root root         81 Feb 28 16:58 gdac.broadinstitute.org_BLCA.RPPA_AnnotateWithGene.Level_3.2014120600.0.0
drwxr-xr-x 2 root root       8192 Feb 28 16:58 gdac.broadinstitute.org_BLCA.Mutation_Packager_Coverage.Level_3.2014120600.0.0
drwxr-xr-x 2 root root       8192 Feb 28 16:49 gdac.broadinstitute.org_BLCA.Mutation_Packager_Calls.Level_3.2014120600.0.0
drwxr-xr-x 2 root root       4096 Feb 28 16:48 gdac.broadinstitute.org_BLCA.mRNAseq_Preprocess.Level_3.2014120600.0.0
drwxr-xr-x 2 root root        154 Feb 28 16:48 gdac.broadinstitute.org_BLCA.miRseq_Mature_Preprocess.Level_3.2014120600.0.0
drwxr-xr-x 2 root root       4096 Feb 28 16:48 gdac.broadinstitute.org_BLCA.Methylation_Preprocess.Level_3.2014120600.0.0
drwxr-xr-x 2 root root        148 Feb 28 16:48 gdac.broadinstitute.org_BLCA.Merge_snp__genome_wide_snp_6__broad_mit_edu__Level_3__segmented_scna_minus_germline_cnv_hg19__seg.Level_3.2014120600.0.0
drwxr-xr-x 2 root root        137 Feb 28 16:48 gdac.broadinstitute.org_BLCA.Merge_rnaseqv2__illuminahiseq_rnaseqv2__unc_edu__Level_3__RSEM_genes_normalized__data.Level_3.2014120600.0.0
drwxr-xr-x 2 root root        136 Feb 28 16:48 gdac.broadinstitute.org_BLCA.Merge_mirnaseq__illuminahiseq_mirnaseq__bcgsc_ca__Level_3__miR_gene_expression__data.Level_3.2014120600.0.0
drwxr-xr-x 2 root root        153 Feb 28 16:46 gdac.broadinstitute.org_BLCA.Merge_methylation__humanmethylation450__jhu_usc_edu__Level_3__within_bioassay_data_set_function__data.Level_3.2014120600.0.0
drwxr-xr-x 2 root root       4096 Feb 28 16:45 gdac.broadinstitute.org_BLCA.Merge_Clinical.Level_1.2014120600.0.0
drwxr-xr-x 2 root root         71 Feb 28 16:45 gdac.broadinstitute.org_BLCA.Clinical_Pick_Tier1.Level_4.2014120600.0.0
-rw-r--r-- 1 root root        153 Feb 28 16:45 head.qsub
-rw-r--r-- 1 root root  352438603 Feb 27 16:41 gdac.broadinstitute.org_BLCA.mRNAseq_Preprocess.Level_3.2014120600.0.0.tar.gz
-rw-r--r-- 1 root root  222593542 Feb 27 16:33 gdac.broadinstitute.org_BLCA.Methylation_Preprocess.Level_3.2014120600.0.0.tar.gz
-rw-r--r-- 1 root root    1521018 Feb 27 16:18 gdac.broadinstitute.org_BLCA.miRseq_Mature_Preprocess.Level_3.2014120600.0.0.tar.gz
-rw-r--r-- 1 root root 1556335569 Feb  4 14:38 gdac.broadinstitute.org_BLCA.Merge_methylation__humanmethylation450__jhu_usc_edu__Level_3__within_bioassay_data_set_function__data.Level_3.2014120600.0.0.tar.gz
-rw-r--r-- 1 root root  408379123 Feb  4 01:07 gdac.broadinstitute.org_BLCA.Mutation_Packager_Coverage.Level_3.2014120600.0.0.tar.gz
-rw-r--r-- 1 root root   34404011 Feb  2 12:02 gdac.broadinstitute.org_BLCA.Merge_rnaseqv2__illuminahiseq_rnaseqv2__unc_edu__Level_3__RSEM_genes_normalized__data.Level_3.2014120600.0.0.tar.gz
-rw-r--r-- 1 root root   10522991 Feb  2 11:46 gdac.broadinstitute.org_BLCA.Mutation_Packager_Calls.Level_3.2014120600.0.0.tar.gz
-rw-r--r-- 1 root root     382957 Feb  2 11:39 gdac.broadinstitute.org_BLCA.RPPA_AnnotateWithGene.Level_3.2014120600.0.0.tar.gz
-rw-r--r-- 1 root root    1809887 Feb  2 11:39 gdac.broadinstitute.org_BLCA.Merge_snp__genome_wide_snp_6__broad_mit_edu__Level_3__segmented_scna_minus_germline_cnv_hg19__seg.Level_3.2014120600.0.0.tar.gz
-rw-r--r-- 1 root root    1668236 Feb  2 11:36 gdac.broadinstitute.org_BLCA.Merge_mirnaseq__illuminahiseq_mirnaseq__bcgsc_ca__Level_3__miR_gene_expression__data.Level_3.2014120600.0.0.tar.gz
-rw-r--r-- 1 root root     542772 Feb  2 11:35 gdac.broadinstitute.org_BLCA.Merge_Clinical.Level_1.2014120600.0.0.tar.gz
-rw-r--r-- 1 root root       7198 Feb  2 11:34 gdac.broadinstitute.org_BLCA.Clinical_Pick_Tier1.Level_4.2014120600.0.0.tar.gz
[zding@console Broad]$
```

Understand data