Materials and Methods

412 OvCa patients in the TCGA database with raw Bam dataset [22] are obtained to generate the expression curve profile (TCGA-23-1023-01 have two bam files), among which 163 patients are sensitive to chemotherapy, and 70 are chemoresistant. Platinum status is defined as resistant if the patient recurred within six months or else sensitive if the platinum free interval is six months or greater, there is no evidence of progression or recurrence, and the follow-up interval is at least six months from the date of last primary platinum treatment defined by TCGA committee [22]. [table, clinical information]. Then we get gene expression[233], methylation[233] dataset of these 233 patients.

TCGA-23-1023-01A-01R-1564-13\_GRCh37-lite\_rnaseq.bam

TCGA-23-1023-01A-02R-1564-13\_GRCh37-lite\_rnaseq.bam

Table 1, clinical information of these 233 samples

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

Compared with patients who are chemosensitive, the chemoresistant patients exhibit relatively poorer overall survival (OS; median, 53.9 vs. 33.8 months; p,0.0001) and progression-free survival (PFS; median, 25.8 vs. 9.3 months; p,0.0001; Figure S3). Other characteristics of these 253 patients are listed in Table 1. The average age at diagnosis is 61.7 years (range, 38.0 to 84.7 years) for the chemoresistant group and 59.1 years (range, 30.5 to 87.5 years) for the chemosensitive group. Up to 84% of the chemosensitive patients show the symptom of recurrent diseases in contrast to 100% of relapse for the chemoresistant patients. 232 among the 253 samples with expression data serve as the TCGA training set

Method,

Allele specific expression

Raw bam files were used to SNP calling with samtools pipline. Firstly, duplication removing with

then

required Phred score >=20 and additionally 20<= number of reads<= 100 for the SNP position. Also excluded positions with indels and >1/4 at the end of a read.

1. Quantitative switch: As I mentioned, we wanted heterozygous SNPs in the treated sample (e.g. A and T allele present at ~50:50) that were homozygous in the control sample (only A or T allele present).
2. Qualitative switch: from one allele to another allele

Disscussion

align reads against a reference genome with IUB ambiguous NA codes. This should remove any allelic bias, you just need to code SNPs into the genome using the IUB codes

Main Statistic of Bam files

<http://www.alexaplatform.org/alexa_seq/5FU/HS04391.htm>

Statistics and figures for sequencing library: HS04391 (Mip101)

Summary of input lanes for this library

Summary of sequence statistics for each lane

Summary of read assignments by assignment class

Summary of mapping results for each lane

Summary of average coverage values by feature type

Summary of expressed events by feature type

Estimates of signal-to-noise ratio

Estimates of intronic and intergenic noise levels (95th percentiles of silent intron and intergenic regions)

Summary of library complexity - estimated by tag redundancy per million reads and compared to other libraries

Distribution relative positions of reads mapping within known transcripts (i.e. position bias test)

Distribution % of gene bases covered for each expressed gene (at various minimum depth levels)

Distribution of log2 raw expression values for each feature type

Distribution of log2 normalized expression values for each feature type

Density scatter plot of exon region versus gene expression values

Density scatter plot of silent intron region versus gene expression values

Distribution of gene-by-gene expression cutoff values

Histograms of expression values for each feature type

Percentiles plot for expression of exon regions, silent intron regions and silent intergenic regions

Cumulative distribution of mean Phred score for all reads