Covariate 1 -Group	Covariate 2 -Stage	Covariate 3 -Window	Covariate 4 -Sample

```
                                    >20  ────────→  Sample 1
                    Early  ──────  10 to 20
         Tg  ──────                 5 to 10
                    16wk  ──────     >20
                                   10 to 20
                    Pre   ──────    0 to 5

                    Early  ──────   N 12wk
         WT  ──────
                    16wk  ──────    N 17wk
                                    N 17wk
                    Pre   ──────    N 19wk   ──────  Sample N
```

1

In each window, it includes different samples. And Window is a more detailed classification of Stage, while stage is a more detailed classification of group. They are dependent, not like other conditions that covariates like age, treatment that are independent.

- How does ComBat deal with sample variance in each group?
- If I use window as a covariate, will ComBat eliminate sample variance in each window?
- If I do not want to eliminate biological variance in samples, what covariates I need to choose? Should I only use sample as a covariates? (Failed due to singularity)

2

Choose covariates

Group	Stage	Window	Sample	ComBat result
+				Good
	+			Good
		+		Good
			+	No singularity
+	+			Good
+		+		No singularity
+			+	No singularity

+: means which factor(s) choose as covariate(s)
Good: means ComBat gave result.
No: means ComBat did not give result due to singularity

3

Difference between using a general covariate and detailed classified covariates

- General covariate: **Group**
- Detailed classified covariates: **Group and Stage**

Stage is a more detailed classification of group. So use stage and group gave a more detailed classification of samples.

Hypothesis:

If there is a difference and it reflects true biological variance, it means a general covariate eliminate true variance and I need to include all detailed classified covariates.

What if the most detailed classification covariates cannot work in ComBat, due to singularity?

4

- A more subtle covariates(group, stage) set gave similar but not identical plot result than 1 covariate one(group).
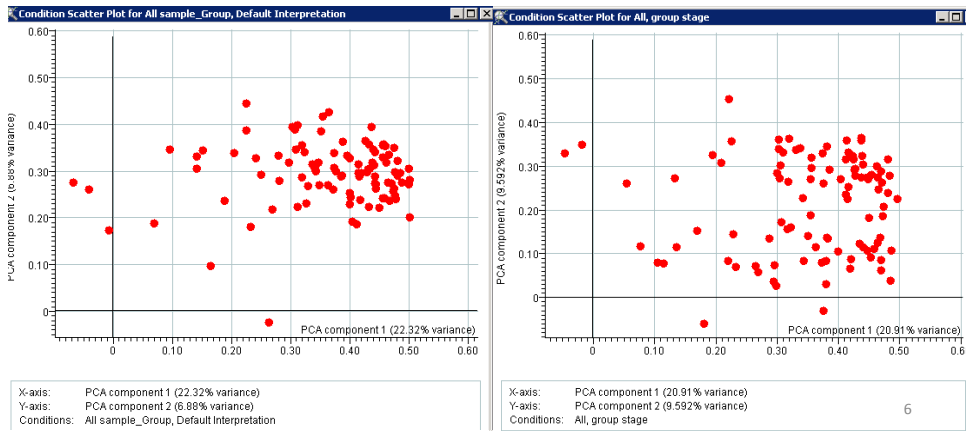
Bandwidth=0.1008   Bandwidth=0.1034
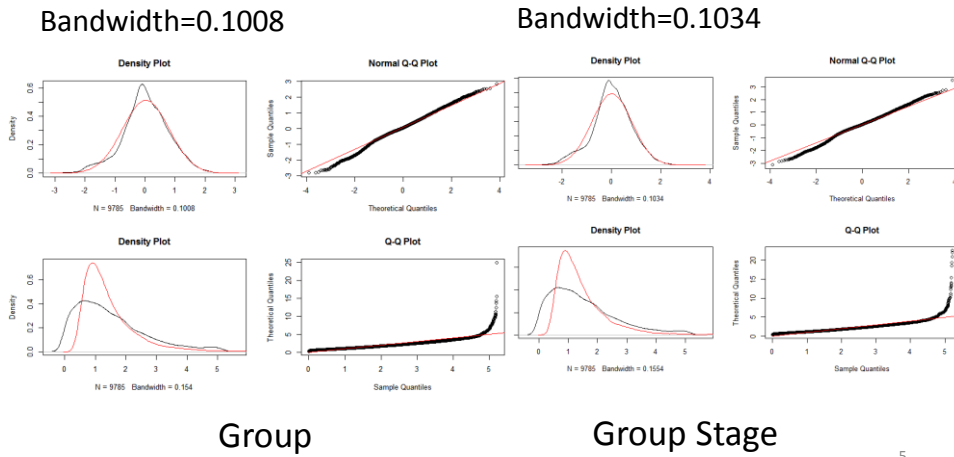
Group   Group Stage

Principle Component analysis gave different result.

group   Group+stage