	CEL	batch	cancer	outcome
18	GSM71037.CEL	1	Cancer	mTCC
20	GSM71039.CEL	1	Cancer	mTCC
22	GSM71041.CEL	1	Cancer	mTCC
28	GSM71047.CEL	1	Cancer	mTCC
29	GSM71048.CEL	1	Cancer	mTCC
31	GSM71050.CEL	1	Cancer	mTCC
32	GSM71051.CEL	1	Cancer	mTCC
33	GSM71052.CEL	1	Cancer	mTCC
35	GSM71054.CEL	1	Cancer	mTCC
40	GSM71060.CEL	1	Cancer	mTCC
46	GSM71066.CEL	1	Cancer	mTCC
2	GSM71020.CEL	2	Normal	Normal
3	GSM71021.CEL	2	Normal	Normal
7	GSM71025.CEL	2	Normal	Normal
8	GSM71026.CEL	2	Normal	Normal
10	GSM71029.CEL	2	Cancer	sTCC-CIS
12	GSM71031.CEL	2	Cancer	sTCC-CIS
14	GSM71033.CEL	2	Cancer	sTCC-CIS
17	GSM71036.CEL	2	Cancer	sTCC-CIS
21	GSM71040.CEL	2	Cancer	mTCC
23	GSM71042.CEL	2	Cancer	sTCC-CIS
25	GSM71044.CEL	2	Cancer	sTCC-CIS
26	GSM71045.CEL	2	Cancer	sTCC-CIS
30	GSM71049.CEL	2	Cancer	sTCC-CIS
36	GSM71055.CEL	2	Cancer	sTCC-CIS
37	GSM71056.CEL	2	Cancer	sTCC-CIS
38	GSM71058.CEL	2	Cancer	sTCC-CIS
39	GSM71059.CEL	2	Cancer	sTCC-CIS
44	GSM71064.CEL	2	Cancer	sTCC-CIS
1	GSM71019.CEL	3	Normal	Normal
4	GSM71022.CEL	3	Normal	Normal
5	GSM71023.CEL	3	Normal	Normal
6	GSM71024.CEL	3	Normal	Normal
49	GSM71069.CEL	4	Biopsy	Biopsy
50	GSM71070.CEL	4	Biopsy	Biopsy
55	GSM71075.CEL	4	Biopsy	Biopsy
56	GSM71076.CEL	4	Biopsy	Biopsy

57	GSM71077.CEL	4	Biopsy	Biopsy
9	GSM71028.CEL	5	Cancer	sTCC+CIS
11	GSM71030.CEL	5	Cancer	sTCC-CIS
13	GSM71032.CEL	5	Cancer	sTCC+CIS
15	GSM71034.CEL	5	Cancer	sTCC+CIS
16	GSM71035.CEL	5	Cancer	sTCC+CIS
19	GSM71038.CEL	5	Cancer	sTCC+CIS
24	GSM71043.CEL	5	Cancer	sTCC+CIS
27	GSM71046.CEL	5	Cancer	sTCC+CIS
34	GSM71053.CEL	5	Cancer	sTCC+CIS
41	GSM71061.CEL	5	Cancer	sTCC+CIS
42	GSM71062.CEL	5	Cancer	sTCC+CIS
43	GSM71063.CEL	5	Cancer	sTCC+CIS
45	GSM71065.CEL	5	Cancer	sTCC-CIS
47	GSM71067.CEL	5	Cancer	sTCC-CIS
48	GSM71068.CEL	5	Cancer	sTCC+CIS
51	GSM71071.CEL	5	Biopsy	Biopsy
52	GSM71072.CEL	5	Biopsy	Biopsy
53	GSM71073.CEL	5	Biopsy	Biopsy
54	GSM71074.CEL	5	Biopsy	Biopsy

How are biological variables and other variables related to study design are distributed among those 5 batches? Explain what could be a problem.

- Individual batches are not diverse enough. Some (batch 1 , 3 and 4) are fully homogenous. Even in batch 2, which has cancer samples with some normal samples mixed in, all outcomes have the same cancer type bar one. The problem we will encounter is that we won't be able to differentiate samples by their biological properties if we divide them by batches.