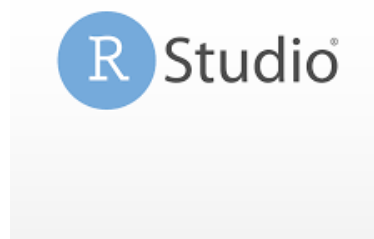
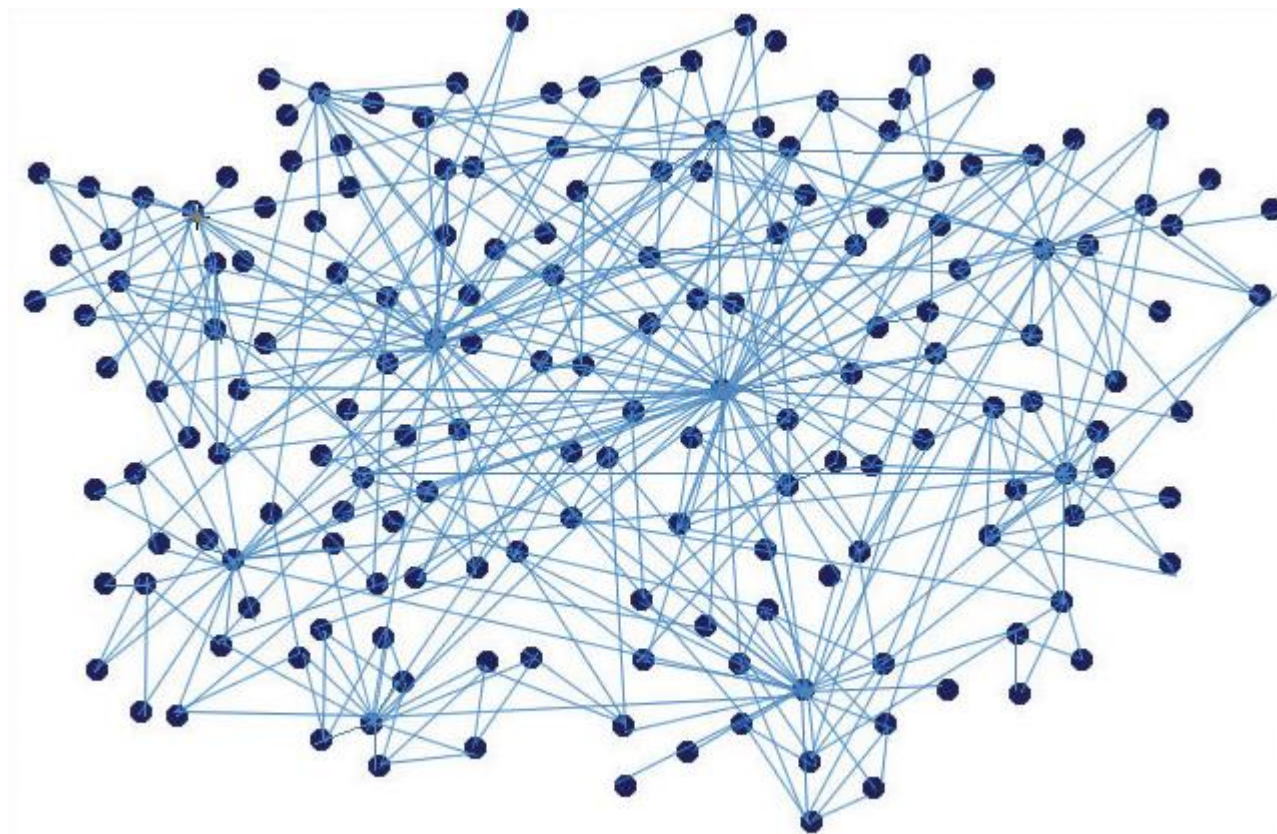


INTRODUZIONE AL DATA MINING

Analisi approfondita sui processi biologici di alcuni pazienti affetti da diverse tipologie di tumori



SOMMARIO

1. Presentazione del progetto.....	3
1.1. Cenni biologici.....	3
1.2. Obiettivo.....	3
2. Modalità d’operazione.....	3
3. Risultati.....	10
4. About.....	12

1. Presentazione del progetto

1.1. Cenni biologici

Il progetto ha inizio con un grafo chiamato “meta-pathway”, che fornisce una rappresentazione di tutti i processi biologici che avvengono nella cellula, attraverso le interazioni tra le molecole (geni).

I nodi di questa rete sono i geni e gli archi direzionati rappresentano le interazioni tra i geni.

Esistono due tipi di interazioni: repressione e stimolazione.

Dunque, per ogni arco vi è una etichetta che ci informa sul tipo di interazione che vi è tra i geni, ovvero “repressione” o “stimolazione”.

1.2. Obiettivo

L’obiettivo del progetto, discusso in questa relazione, riguarda l’analisi e l’individuazione di pazienti con tumori simili che hanno gli stessi motivi, ovvero gli stessi processi biologici o parti di processi biologici alterati allo stesso modo tra pazienti con le stesse patologie.

2. Modalità d’operazione

Partendo dai dati di espressione dei geni in cui viene misurata la quantità di gene presente nelle cellule di un paziente, è stato eseguito un algoritmo chiamato

“Mithrill” che è servito a mappare i valori di espressione nei corrispondenti geni della rete ed ha assegnato per ogni gene di ogni paziente un valore di perturbazione, che può essere 0, positivo o negativo. Si annotano, quindi, in “signature_Main_Classification.txt”:

```
signature_Main_Classification.txt
C05140 "cpd:C05172" "cpd:C05212" "cpd:C05259" "cpd:C05261" "cpd:C05263" "cpd:C05265" "cpd:C05267" "cpd:C05269" "cpd:
C05271" "cpd:C05272" "cpd:C05273" "cpd:C05275" "cpd:C05276" "cpd:C05279" "cpd:C05285" "cpd:C05293" "cpd:C05300" "cpd:
C05336" "cpd:C05340" "cpd:C05341" "cpd:C05345" "cpd:C05350" "cpd:C05379" "cpd:C05439" "cpd:C05449" "cpd:C05455" "cpd:
C05460" "cpd:C05465" "cpd:C05466" "cpd:C05470" "cpd:C05472" "cpd:C05474" "cpd:C05478" "cpd:C05480" "cpd:C05490" "cpd:
C05502" "cpd:C05503" "cpd:C05504" "cpd:C05512" "cpd:C05517" "cpd:C05528" "cpd:C05546" "cpd:C05579" "cpd:C05582" "cpd:
C05584" "cpd:C05594" "cpd:C05598" "cpd:C05606" "cpd:C05635" "cpd:C05637" "cpd:C05639" "cpd:C05640" "cpd:C05642" "cpd:
C05643" "cpd:C05645" "cpd:C05647" "cpd:C05660" "cpd:C05675" "cpd:C05686" "cpd:C05703" "cpd:C05711" "cpd:C05720" "cpd:
C05744" "cpd:C05750" "cpd:C05753" "cpd:C05762" "cpd:C05768" "cpd:C05791" "cpd:C05803" "cpd:C05804" "cpd:C05805" "cpd:
C05813" "cpd:C05814" "cpd:C05815" "cpd:C05823" "cpd:C05828" "cpd:C05842" "cpd:C05843" "cpd:C05844" "cpd:C05848" "cpd:
C05852" "cpd:C05859" "cpd:C05924" "cpd:C05951" "cpd:C05956" "cpd:C05959" "cpd:C05965" "cpd:C05974" "cpd:C05980" "cpd:
C05981" "cpd:C05998" "cpd:C06054" "cpd:C06055" "cpd:C06112" "cpd:C06113" "cpd:C06114" "cpd:C06124" "cpd:C06125" "cpd:
C06127" "cpd:C06178" "cpd:C06196" "cpd:C06212" "cpd:C06213" "cpd:C06250" "cpd:C06329" "cpd:C06423" "cpd:C06424" "cpd:
C06425" "cpd:C06426" "cpd:C06427" "cpd:C06428" "cpd:C06429" "cpd:C06482" "cpd:C06506" "cpd:C06508" "cpd:C06611" "cpd:
C06612" "cpd:C06754" "cpd:C07080" "cpd:C07496" "cpd:C07585" "cpd:C07644" "cpd:C07646" "cpd:C07712" "cpd:C08270" "cpd:
C11039" "cpd:C11061" "cpd:C11131" "cpd:C11132" "cpd:C11133" "cpd:C11134" "cpd:C11135" "cpd:C11136" "cpd:C11148" "cpd:
C11150" "cpd:C11278" "cpd:C11522" "cpd:C11554" "cpd:C11583" "cpd:C11907" "cpd:C12144" "cpd:C12145" "cpd:C12455" "cpd:
C12834" "cpd:C13050" "cpd:C13630" "cpd:C13713" "cpd:C14748" "cpd:C14749" "cpd:C14772" "cpd:C14773" "cpd:C14775" "cpd:
C14775" "cpd:C14782" "cpd:C14787" "cpd:C14791" "cpd:C14801" "cpd:C14802" "cpd:C14812" "cpd:C14814" "cpd:C14819" "cpd:
C14823" "cpd:C14825" "cpd:C14826" "cpd:C14845" "cpd:C14847" "cpd:C14848" "cpd:C14849" "cpd:C14853" "cpd:C14854" "cpd:
C14855" "cpd:C14856" "cpd:C14861" "cpd:C14863" "cpd:C14864" "cpd:C14866" "cpd:C14868" "cpd:C14869" "cpd:C14871" "cpd:
C14874" "cpd:C15492" "cpd:C15493" "cpd:C15606" "cpd:C15645" "cpd:C15646" "cpd:C15647" "cpd:C15670" "cpd:C15778" "cpd:
C15804" "cpd:C15805" "cpd:C15806" "cpd:C15812" "cpd:C15915" "cpd:C15977" "cpd:C15979" "cpd:C16163" "cpd:C16164" "cpd:
C16169" "cpd:C16173" "cpd:C16219" "cpd:C16237" "cpd:C16254" "cpd:C16255" "cpd:C16300" "cpd:C16328" "cpd:C16330" "cpd:
C16332" "cpd:C16334" "cpd:C16336" "cpd:C16338" "cpd:C16339" "cpd:C16355" "cpd:C16356" "cpd:C16359" "cpd:C16360" "cpd:
C16365" "cpd:C16344" "cpd:C16345" "cpd:C16347" "cpd:C16348" "cpd:C16349" "cpd:C16353" "cpd:C16356" "cpd:C16359" "cpd:C16360" "cpd:
C16360" "cpd:C16361" "cpd:C16365" "cpd:C16376" "cpd:C16377" "cpd:C16378" "cpd:C16382" "cpd:C16384" "cpd:C16391" "cpd:
C16396" "cpd:C16602" "cpd:C16604" "cpd:C16607" "cpd:C16610" "cpd:C16613" "cpd:C16615" "cpd:C16620" "cpd:C16622" "cpd:
C16632" "cpd:C16634" "cpd:C16643" "cpd:C16651" "cpd:C16660" "cpd:C16662" "cpd:C16664" "cpd:C16677" "cpd:C16679" "cpd:
C16680" "cpd:C17331" "cpd:C17332" "cpd:C17337" "cpd:C17432" "cpd:C17559" "cpd:C17560" "cpd:C17561" "cpd:C17562" "cpd:
C17938" "cpd:C18038" "cpd:C18042" "cpd:C18043" "cpd:C18044" "cpd:C18045" "cpd:C18075" "cpd:C18231" "cpd:C18239" "cpd:
C18872" "cpd:C18904" "cpd:C19559" "cpd:C19562" "cpd:C19563" "cpd:C19566" "cpd:C19577" "cpd:C19580" "cpd:C19585" "cpd:
C19588" "cpd:C19594" "cpd:C19595" "cpd:C19607" "cpd:C19692" "cpd:C20238" "cpd:C20239" "cpd:C20372" "cpd:C20376" "cpd:
C20690" "cpd:C20775" "cpd:C20825" "cpd:C21016" "cpd:C21106" "gl:G00018" "gl:G00019" "gl:G00022" "gl:G00024" "gl:G
00027" "gl:G00028" "gl:G00029" "gl:G00031" "gl:G00032" "gl:G00035" "gl:G00040" "gl:G00043" "gl:G00045" "gl:G
00046" "gl:G00048" "gl:G00052" "gl:G00054" "gl:G00056" "gl:G00059" "gl:G00063" "gl:G00064" "gl:G00065" "gl:G
00068" "gl:G00072" "gl:G00075" "gl:G00076" "gl:G00077" "gl:G00079" "gl:G00082" "gl:G00083" "gl:G00086" "gl:G
00088" "gl:G00090" "gl:G00098" "gl:G00099" "gl:G00104" "gl:G00112" "gl:G00082" "gl:G00083" "gl:G00086" "gl:G
00127" "gl:G00128" "gl:G00129" "gl:G00147" "gl:G00160" "gl:G00112" "gl:G00117" "gl:G00120" "gl:G00126" "gl:G
02632" "gl:G12306" "gl:G13030" "gl:G13044" "gl:G13045" "gl:G13045" "gl:G00711" "gl:G00872" "gl:G01391" "gl:G
"TCGA-OR-ASJ1-01A" -4.60416350743688 -0.123044483997715 -2.25591928746909 0
1.56935091596381 -0.269834006604705 0 0.82890468957192 -1.8466101236465
3.18104811193777 -0.041917549824319 -2.40556614114083 0 0.120320317564146 0.0889524412154558
0.0889524412154558 0.0889524412154558 0 1.9080947177336 0.0889524412154558 -2.00557696820923
0 -0.158766947478585
1.24948344389716 -2.04354986422131 -0.731350836311607 -2.35212005821196 -4.72087489722086 -0.597834358103402
-2.44832268830225 0 -4.44494512032384 -0.20393989475689 0 -4.67997279403049 -1.06639279276564
0.562092973042496 -1.38312033182763 3.70248303600909 0 1.3355110744327
0.812490160033862 -0.273403027576948 -0.273403027576948 0.0606963208311595 0 -3.80596856527377
8.94526920191712 0 -0.212789141777549 -0.0758112735650356 -0.530697170292624 0 -2.050770548105 0
0.115391094106203 -1.06943826296777 0 -0.726978788860302 0 -2.050770548105 0
0.0889524412154558 -10.2107356576967 -4.81305295271125 -0.483657064183074 -0.212789141777549 -11.8943721532663
-2.16521000598107 -4.71302182143484 0.236655869403826 0 -0.212789141777549 -11.8943721532663
0.236655869403826 -0.288756630437278 -0.23433709180578 0.856762485054752 0 -0.425402953855909
0.0310739902409302 -0.956704738578574 -9.04481555089694 -15.3064294270365 -2.93935701107725
```

Tramite l’algoritmo “./Solution.cpp” è stato effettuato un controllo partendo dal file “./MetaPathway/meta_nodes.txt” che contiene tutti i geni d’interesse per l’analisi i quali sono stati ricercati successivamente nel file “signature_Main_Classification.txt”, generando un file per ogni paziente (in ./reti-pazienti/risultati-map.txt). In ognuno di essi è stato, quindi, annotato ogni valore di perturbazione corrispondente ad ogni gene interessato preso in esame. Il risultato è la “meta-pathway” di ogni paziente. In seguito, si è proceduto con

la sostituzione dei valori di perturbazione dei meta-pathway di ogni paziente con le relative sigle:

- PP: se il valore di perturbazione è positivo;
- UP: se il valore di perturbazione è 0;
- NP: se il valore di perturbazione è negativo;

Si ottengono, quindi, i seguenti risultati in “./reti-pazienti/risultati-solution.txt”:

risultati-solution				
Nome	Data di modifica	Dimensioni	Tipo	
TCGA-02-0047-01A.txt	11 aprile 2019 20:12	573 KB	Solo testo	
TCGA-02-0055-01A.txt	11 aprile 2019 20:10	573 KB	Solo testo	
TCGA-02-2483-01A.txt	11 aprile 2019 20:09	573 KB	Solo testo	
TCGA-02-2485-01A.txt	11 aprile 2019 20:08	573 KB	Solo testo	
TCGA-02-2486-01A.txt	11 aprile 2019 20:10	573 KB	Solo testo	
TCGA-2A-A8VL-01A.txt	11 aprile 2019 21:14	573 KB	Solo testo	
TCGA-2A-A8VO-01A.txt	11 aprile 2019 21:21	573 KB	Solo testo	
TCGA-2A-A8VT-01A.txt	11 aprile 2019 21:14	573 KB	Solo testo	
TCGA-2A-A8VV-01A.txt	11 aprile 2019 21:14	573 KB	Solo testo	
TCGA-2A-A8W1-01A.txt	11 aprile 2019 21:18	573 KB	Solo testo	
TCGA-2A-A8W3-01A.txt	11 aprile 2019 21:18	573 KB	Solo testo	
TCGA-2E-A9G8-01A.txt	11 aprile 2019 22:19	573 KB	Solo testo	
TCGA-2F-A9KO-01A.txt	11 aprile 2019 19:11	573 KB	Solo testo	
TCGA-2F-A9KP-01A.txt	11 aprile 2019 19:14	573 KB	Solo testo	
TCGA-2F-A9KQ-01A.txt	11 aprile 2019 19:15	573 KB	Solo testo	
TCGA-2F-A9KR-01A.txt	11 aprile 2019 19:15	573 KB	Solo testo	
TCGA-2F-A9KT-01A.txt	11 aprile 2019 19:15	573 KB	Solo testo	
TCGA-2F-A9KW-01A.txt	11 aprile 2019 19:14	573 KB	Solo testo	
TCGA-2G-AAEW-01A.txt	11 aprile 2019 21:57	573 KB	Solo testo	
TCGA-2G-AAEX-01A.txt	11 aprile 2019 21:57	573 KB	Solo testo	
TCGA-2G-AAF1-01A.txt	11 aprile 2019 21:57	573 KB	Solo testo	
TCGA-2G-AAF4-01A.txt	11 aprile 2019 21:57	573 KB	Solo testo	
TCGA-2G-AAF6-01A.txt	11 aprile 2019 21:57	573 KB	Solo testo	
TCGA-2G-AAF8-01A.txt	11 aprile 2019 21:57	573 KB	Solo testo	
TCGA-2G-AAFG-01A.txt	11 aprile 2019 21:57	573 KB	Solo testo	
TCGA-2G-AAFG-05A.txt	11 aprile 2019 21:57	573 KB	Solo testo	
TCGA-2G-AAFH-01A.txt	11 aprile 2019 21:57	573 KB	Solo testo	

La terza fase del progetto ha previsto l'uso di due algoritmi "Namecopy.cpp" e "MultiMotif.cpp". Il primo è servito ad estrapolare tutti i nomi dei pazienti presi in osservazione annotandoli in "nametocolumn.txt"; il secondo, invece, per ogni paziente annotato in "nametocolumn.txt", ha fatto partire da riga di comando il software *Multimotif*,






























un software in grado di calcolare i motivi per i meta-pathway di ogni paziente, trovando i sottografi etichettati più ricorrenti nella rete. I risultati si trovano in `"/MultiMotif/motif"`.

La quarta fase prevede la generazione di una matrice finale che conterrà per colonna tutti i motivi che verranno estrapolati tramite l'algoritmo "Create-Map-Motif_Nodes__Motif_edges.cpp" e per righe tutti i pazienti con i rispettivi valori per quei motivi, 0 se non presenti. Il primo step che viene eseguito è l'estrapolazione di tutti i tipi di motivi presenti tra tutti i risultati di Multimotif che vengono annotati in `"/MultiMotif/map-all-Motif_Nodes-Motif_edges.txt"`. Poi si procede con la generazione della matrice finale (`"/Matrix/Analysis_Matrix.txt"`), illustrata sopra, tramite l'algoritmo "Final_matrix.cpp" che fa uso di tutti i motivi estrapolati in precedenza e annotati in `"/MultiMotif/map-all-Motif_Nodes-Motif_edges.txt"`.

	"NP, NP, NP-(0,1,activation)"	(1,0,activation)"	(2,1,activation)"	(2,0,activation)"	"NP, NP, NP-(0,1,activation)"	(1,0,activation)"	(2,1,inhibition)"	(2,0,activation)"	"NP, NP, NP-																							
1	"TCGA-OR-A5J1-01A"	1070	1	132	11	6	583	71	7	1409	3	4164	705	1328	220	7	1	173	1	0	48	35	38955	4037	48	881	3	4326	406	1		
2	"TCGA-OR-A5J2-01A"	1070	1	66	17	6	547	69	7	1393	2	4285	705	1328	220	7	1	134	1	0	42	47	37885	3865	57	837	4	4233	240	9		
3	"TCGA-OR-A5J3-01A"	80	1	0	17	0	129	65	7	73	2	2303	45	8	0	7	1	152	1	0	41	47	36109	1903	50	177	4	3939	108	12	0	43804
4	"TCGA-OR-A5J5-01A"	215	0	15	10	0	162	55	7	192	0	2414	105	128	20	6	0	189	0	33	41	35705	2153	40	237	3	4459	135	12	9	42935	
5	"TCGA-OR-A5J7-01A"	1070	1	66	7	0	568	69	7	1393	2	4162	705	1328	220	1	1	160	1	0	28	42	37482	3856	12	835	0	4037	209	0		
6	"TCGA-OR-A5J8-01A"	191	0	15	11	0	157	53	4	208	6	1965	106	128	20	6	0	188	0	25	26	21581	1995	34	280	3	2999	151	15	8	32127	
7	"TCGA-OR-A5J9-01A"	1070	1	66	17	6	569	65	7	1393	2	4278	705	1328	220	7	1	148	1	0	42	45	37818	3859	52	834	4	4134	231	1		
8	"TCGA-OR-A5JA-01A"	1070	1	66	11	0	583	71	7	1409	3	4397	705	1328	220	7	1	175	1	0	37	53	39077	3995	41	881	3	4563	274	1		
9	"TCGA-OR-A5JB-01A"	1070	0	66	8	0	564	67	7	1414	6	3897	704	1328	220	6	0	204	0	0	17	5	26831	3927	39	906	4	3277	286	1		
10	"TCGA-OR-A5JC-01A"	215	0	15	19	6	190	70	7	214	6	2430	106	128	20	6	0	159	0	47	17	37250	2382	59	308	4	3757	187	15	8	45120	
11	"TCGA-OR-A5JF-01A"	215	0	15	17	6	172	63	7	196	0	2466	104	128	20	6	0	148	0	41	47	36487	2189	57	254	4	3938	143	12	8	44123	
12	"TCGA-OR-A5JG-01A"	1070	1	66	11	6	585	78	7	1415	6	4417	705	1328	220	7	1	154	1	0	46	36	39874	4123	48	908	3	4370	292	1		
13	"TCGA-OR-A5JJ-01A"	199	0	15	10	6	170	71	7	196	4	2363	104	128	20	6	0	161	0	41	45	34548	2069	44	252	3	4237	140	12	8	42841	
14	"TCGA-OR-A5JM-01A"	1070	0	66	19	0	590	70	7	1414	6	4429	704	1328	220	6	0	155	0	41	53	39945	4092	49	908	4	4473	289	1			
15	"TCGA-OR-A5JP-01A"	215	0	15	11	4	184	70	7	214	6	2726	107	128	20	6	0	159	0	45	53	38018	2394	45	308	3	4268	189	15	9	47345	
16	"TCGA-OR-A5JS-01A"	215	0	15	11	0	188	74	7	214	6	2756	104	128	20	6	0	183	0	41	50	39012	2392	41	308	3	4444	187	15	8	47997	
17	"TCGA-OR-A5JT-01A"	1070	1	66	17	0	573	69</																								

Prima di arrivare alla quinta fase, ovvero la clusterizzazione dei risultati, la matrice finale (“./Matrix/Analysis_Matrix.txt”) viene divisa, tramite l’algoritmo “Matrix_division.cpp”, per tipologia di tumore, cioè tutti i pazienti affetti dalla stessa tipologia di tumore verranno inseriti in un unico file e ,di conseguenza, si avranno tanti file quanti sono i tumori presi in osservazione. Il risultato di tale divisione è contenuto in ‘./Matrix/type_division/’:

type_division				
Nome	^	Data di modifica	Dimensioni	Tipo
 Matrix-ACC.txt		13 maggio 2019 15:03	460 KB	Solo testo
 Matrix-BLCA.txt		13 maggio 2019 15:03	2 MB	Solo testo
 Matrix-BRCA.txt		13 maggio 2019 15:03	5,1 MB	Solo testo
 Matrix-CESC.txt		13 maggio 2019 15:03	582 KB	Solo testo
 Matrix-CHOL.txt		13 maggio 2019 15:03	262 KB	Solo testo
 Matrix-COAD.txt		13 maggio 2019 15:03	579 KB	Solo testo
 Matrix-DLBC.txt		13 maggio 2019 15:03	111 KB	Solo testo
 Matrix-ESCA.txt		13 maggio 2019 15:03	894 KB	Solo testo
 Matrix-GBM.txt		13 maggio 2019 15:03	860 KB	Solo testo
 Matrix-HNSC.txt		13 maggio 2019 15:03	1,4 MB	Solo testo
 Matrix-KICH.txt		13 maggio 2019 15:03	111 KB	Solo testo
 Matrix-KIRC.txt		13 maggio 2019 15:03	111 KB	Solo testo
 Matrix-KIRP.txt		13 maggio 2019 15:03	848 KB	Solo testo
 Matrix-LAML.txt		13 maggio 2019 15:03	910 KB	Solo testo
 Matrix-LGG.txt		13 maggio 2019 15:03	2,5 MB	Solo testo
 Matrix-LIHC.txt		13 maggio 2019 15:03	129 KB	Solo testo
 Matrix-LUAD.txt		13 maggio 2019 15:03	1,2 MB	Solo testo
 Matrix-LUSC.txt		13 maggio 2019 15:03	111 KB	Solo testo
 Matrix-MESO.txt		13 maggio 2019 15:03	111 KB	Solo testo
 Matrix-OV.txt		13 maggio 2019 15:03	1,5 MB	Solo testo
 Matrix-PAAD.txt		13 maggio 2019 15:03	111 KB	Solo testo
 Matrix-PCPG.txt		13 maggio 2019 15:03	921 KB	Solo testo
 Matrix-PRAD.txt		13 maggio 2019 15:03	1,5 MB	Solo testo
 Matrix-READ.txt		13 maggio 2019 15:03	421 KB	Solo testo
 Matrix-SARC.txt		13 maggio 2019 15:03	1,1 MB	Solo testo
 Matrix-SKCM.txt		13 maggio 2019 15:03	1,5 MB	Solo testo
 Matrix-STAD.txt		13 maggio 2019 15:03	1,9 MB	Solo testo

Successivamente viene applicato un secondo algoritmo, “Type_of_tumor.cpp”, che genera due file contenenti uno il numero di sottotipi di tumore per ogni tipologia di tumore, e l'altro i nomi dei sotto-tipi ,ottenendo per entrambi uno schema chiave-valore. È possibile visionare i file in “./map_subtumor_tumor.txt”, “./number-subtumor_for_tumor.txt”.

Questi serviranno a stabilire l’efficienza del clustering così da capire se è stato di tipo ottimale o meno per ogni tumore. Vale a dire, se un determinato sotto-tipo di tumore è contenuto solamente in un cluster allora la clusterizzazione avrà avuto un buon riscontro, mentre, se un sotto-tipo di tumore si presenta in più cluster dello stesso tumore, allora il cluster non sarà stato totalmente efficiente.

Dunque, l'ultima fase di clusterizzazione avviene tramite un algoritmo `“./Cluster/nometumore/Cluster-nometumore.R”` che effettua un clustering basato su densità, ovvero il DBSCAN, collocando i risultati in ogni cartella relativa ai tumori, ovvero in `“./Cluster/nometumore/nometumore-cluster.txt”`.

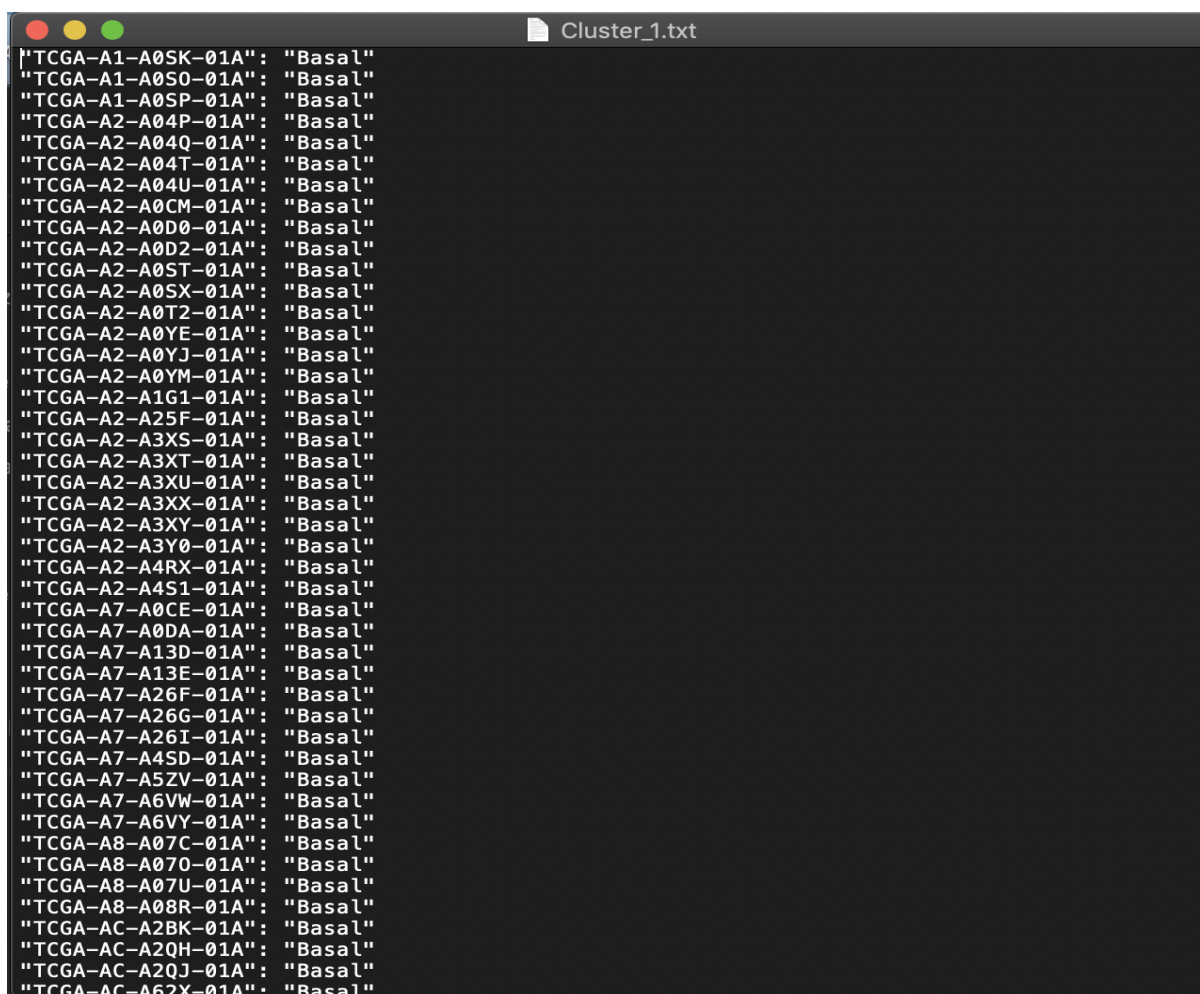
Cluster				
Nome	Data di modifica	Dimensioni	Tipo	
▶ ACC	17 maggio 2019 12:31	--	Cartella	
▶ BLCA	15 maggio 2019 18:23	--	Cartella	
▶ BRCA	17 maggio 2019 12:38	--	Cartella	
▶ CESC	15 maggio 2019 18:33	--	Cartella	
▶ CHOL	15 maggio 2019 18:44	--	Cartella	
▶ COAD	15 maggio 2019 18:47	--	Cartella	
▶ ESCA	15 maggio 2019 20:16	--	Cartella	
▶ GBM	16 maggio 2019 02:48	--	Cartella	
▶ HNSC	16 maggio 2019 02:51	--	Cartella	
▶ KIRP	16 maggio 2019 02:55	--	Cartella	
▶ LAML	16 maggio 2019 02:59	--	Cartella	
▶ LGG	16 maggio 2019 03:02	--	Cartella	
▶ LIHC	16 maggio 2019 11:08	--	Cartella	
▶ LUAD	16 maggio 2019 11:12	--	Cartella	
▶ OV	16 maggio 2019 11:15	--	Cartella	
▶ PCPG	16 maggio 2019 11:18	--	Cartella	
▶ PRAD	16 maggio 2019 11:19	--	Cartella	
▶ READ	16 maggio 2019 11:21	--	Cartella	
▶ SARC	16 maggio 2019 11:23	--	Cartella	
▶ SKCM	16 maggio 2019 11:42	--	Cartella	
▶ STAD	16 maggio 2019 11:33	--	Cartella	
▶ TGCT	16 maggio 2019 11:36	--	Cartella	
▶ THCA	16 maggio 2019 11:41	--	Cartella	
▶ THYM	16 maggio 2019 11:44	--	Cartella	
▶ UCEC	16 maggio 2019 11:45	--	Cartella	
▶ UCS	16 maggio 2019 11:47	--	Cartella	
▶ UVM	16 maggio 2019 11:49	--	Cartella	

ACC				
Nome	Data di modifica	Dimensioni	Tipo	
ACC-cluster.txt	15 maggio 2019 15:03	2 KB	Solo testo	
Cluster_1.txt	17 maggio 2019 12:31	1 KB	Solo testo	
Cluster_2.txt	17 maggio 2019 12:31	598 byte	Solo testo	
Cluster_punti_outlier.txt	17 maggio 2019 12:31	26 byte	Solo testo	
Cluster-ACC.R	15 maggio 2019 15:03	402 byte	Rez Source	

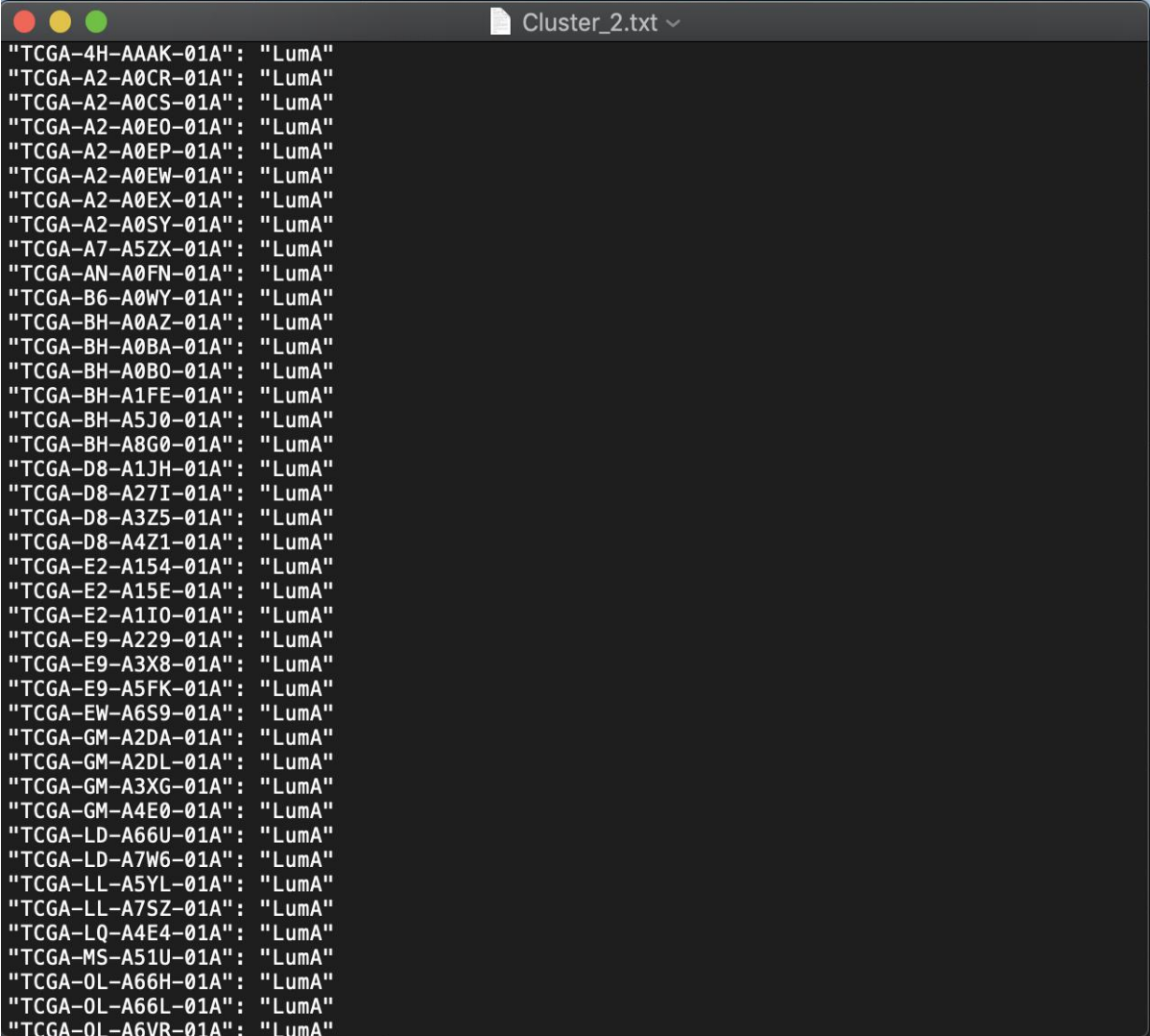
Infine, tramite l'algoritmo "cluster_classification.cpp", sono stati suddivisi i pazienti di uno stesso tumore (già clusterizzati) per il numero di cluster calcolato, ovvero vi sarà un file per ogni cluster calcolato di ogni tumore e i pazienti verranno inseriti nel cluster che l'algoritmo "Cluster-nometumore.R" ha calcolato per ognuno di essi.

3. Risultati

I risultati ottenuti per cluster variano dai più ottimali a quelli meno ottimali. Tra tutti i tumori emergono il BLCA e il BRCA, in cui ogni sotto tipo di tumore è contenuto in un solo cluster e quindi, come definito sopra, questo indica che è stato effettuato un buon processo di clusterizzazione:



```
Cluster_1.txt
"TCGA-A1-A0SK-01A": "Basal"
"TCGA-A1-A0S0-01A": "Basal"
"TCGA-A1-A0SP-01A": "Basal"
"TCGA-A2-A04P-01A": "Basal"
"TCGA-A2-A04Q-01A": "Basal"
"TCGA-A2-A04T-01A": "Basal"
"TCGA-A2-A04U-01A": "Basal"
"TCGA-A2-A0CM-01A": "Basal"
"TCGA-A2-A0D0-01A": "Basal"
"TCGA-A2-A0D2-01A": "Basal"
"TCGA-A2-A0ST-01A": "Basal"
"TCGA-A2-A0SX-01A": "Basal"
"TCGA-A2-A0T2-01A": "Basal"
"TCGA-A2-A0YE-01A": "Basal"
"TCGA-A2-A0YJ-01A": "Basal"
"TCGA-A2-A0YM-01A": "Basal"
"TCGA-A2-A1G1-01A": "Basal"
"TCGA-A2-A25F-01A": "Basal"
"TCGA-A2-A3XS-01A": "Basal"
"TCGA-A2-A3XT-01A": "Basal"
"TCGA-A2-A3XU-01A": "Basal"
"TCGA-A2-A3XX-01A": "Basal"
"TCGA-A2-A3XY-01A": "Basal"
"TCGA-A2-A3Y0-01A": "Basal"
"TCGA-A2-A4RX-01A": "Basal"
"TCGA-A2-A4S1-01A": "Basal"
"TCGA-A7-A0CE-01A": "Basal"
"TCGA-A7-A0DA-01A": "Basal"
"TCGA-A7-A13D-01A": "Basal"
"TCGA-A7-A13E-01A": "Basal"
"TCGA-A7-A26F-01A": "Basal"
"TCGA-A7-A26G-01A": "Basal"
"TCGA-A7-A26I-01A": "Basal"
"TCGA-A7-A4SD-01A": "Basal"
"TCGA-A7-A5ZV-01A": "Basal"
"TCGA-A7-A6VW-01A": "Basal"
"TCGA-A7-A6VY-01A": "Basal"
"TCGA-A8-A07C-01A": "Basal"
"TCGA-A8-A070-01A": "Basal"
"TCGA-A8-A07U-01A": "Basal"
"TCGA-A8-A08R-01A": "Basal"
"TCGA-AC-A2BK-01A": "Basal"
"TCGA-AC-A2QH-01A": "Basal"
"TCGA-AC-A2QJ-01A": "Basal"
"TCGA-AC-A62X-01A": "Basal"
```



```
"TCGA-4H-AAAK-01A": "LumA"  
"TCGA-A2-A0CR-01A": "LumA"  
"TCGA-A2-A0CS-01A": "LumA"  
"TCGA-A2-A0E0-01A": "LumA"  
"TCGA-A2-A0EP-01A": "LumA"  
"TCGA-A2-A0EW-01A": "LumA"  
"TCGA-A2-A0EX-01A": "LumA"  
"TCGA-A2-A0SY-01A": "LumA"  
"TCGA-A7-A5ZX-01A": "LumA"  
"TCGA-AN-A0FN-01A": "LumA"  
"TCGA-B6-A0WY-01A": "LumA"  
"TCGA-BH-A0AZ-01A": "LumA"  
"TCGA-BH-A0BA-01A": "LumA"  
"TCGA-BH-A0B0-01A": "LumA"  
"TCGA-BH-A1FE-01A": "LumA"  
"TCGA-BH-A5J0-01A": "LumA"  
"TCGA-BH-A8G0-01A": "LumA"  
"TCGA-D8-A1JH-01A": "LumA"  
"TCGA-D8-A27I-01A": "LumA"  
"TCGA-D8-A3Z5-01A": "LumA"  
"TCGA-D8-A4Z1-01A": "LumA"  
"TCGA-E2-A154-01A": "LumA"  
"TCGA-E2-A15E-01A": "LumA"  
"TCGA-E2-A1I0-01A": "LumA"  
"TCGA-E9-A229-01A": "LumA"  
"TCGA-E9-A3X8-01A": "LumA"  
"TCGA-E9-A5FK-01A": "LumA"  
"TCGA-EW-A6S9-01A": "LumA"  
"TCGA-GM-A2DA-01A": "LumA"  
"TCGA-GM-A2DL-01A": "LumA"  
"TCGA-GM-A3XG-01A": "LumA"  
"TCGA-GM-A4E0-01A": "LumA"  
"TCGA-LD-A66U-01A": "LumA"  
"TCGA-LD-A7W6-01A": "LumA"  
"TCGA-LL-A5YL-01A": "LumA"  
"TCGA-LL-A7SZ-01A": "LumA"  
"TCGA-LQ-A4E4-01A": "LumA"  
"TCGA-MS-A51U-01A": "LumA"  
"TCGA-OL-A66H-01A": "LumA"  
"TCGA-OL-A66L-01A": "LumA"  
"TCGA-OL-A6VR-01A": "LumA"
```

4. About

Giovanni Martucci, X81000188.

Studiante presso l'Università di Matematica ed Informatica di Catania, appassionato di Digital Forensics e Data-Mining.