esiCancer: Evolutionary *in Silico* Cancer Simulator Documentation

Minussi D.C., Henz B., Oliveira M.S.,Filippi-Chiela E.C., Oliveira M.M., Lenz G.

Contents

1 Description

esiCancer is a cell-autonomous software tool to model evolution of cancer. It is a fully-customizable tool, in which real genomic information of tumors can be used to construct a biallelic genome and stochastic events, such as mutations, indels and other alterations can be applied to this genome. Each event can produce several alterations to the probability of division, death, event rate per division, and/or maximal divisions, which impact the population of cells over time.

esiCancer has a user friendly interface and produces .cvs outputs that can be easily analyzed by computer professionals or amateurs alike.

2 Getting Start

2.1 Compiling the code

Source code is written in C++, using Qt library for interface design and it available in `https://github.com/bernardohenz/esiCancer`. For compiling the code you will need to: (1) have a compiler installed (we suggest GCC or MinGW, but any compiler should work fine); (2) have the QT library, which can be downloaded from the QT website. We strongly recommend the use of QtCreator (installed with QT), as we provide the .pro project file. Any problem compiling the code, please let me know.

2.2 Installing esiCancer

The compiled version of esiCancer can be download from `http://www.ufrgs.br/labsinal/esiCancer/index.html`, in which is available for macOs, windows and linux. For first time users, it is necessary open the program as administrator.

3 esiCancer Interface

3.1 Interface Parameters
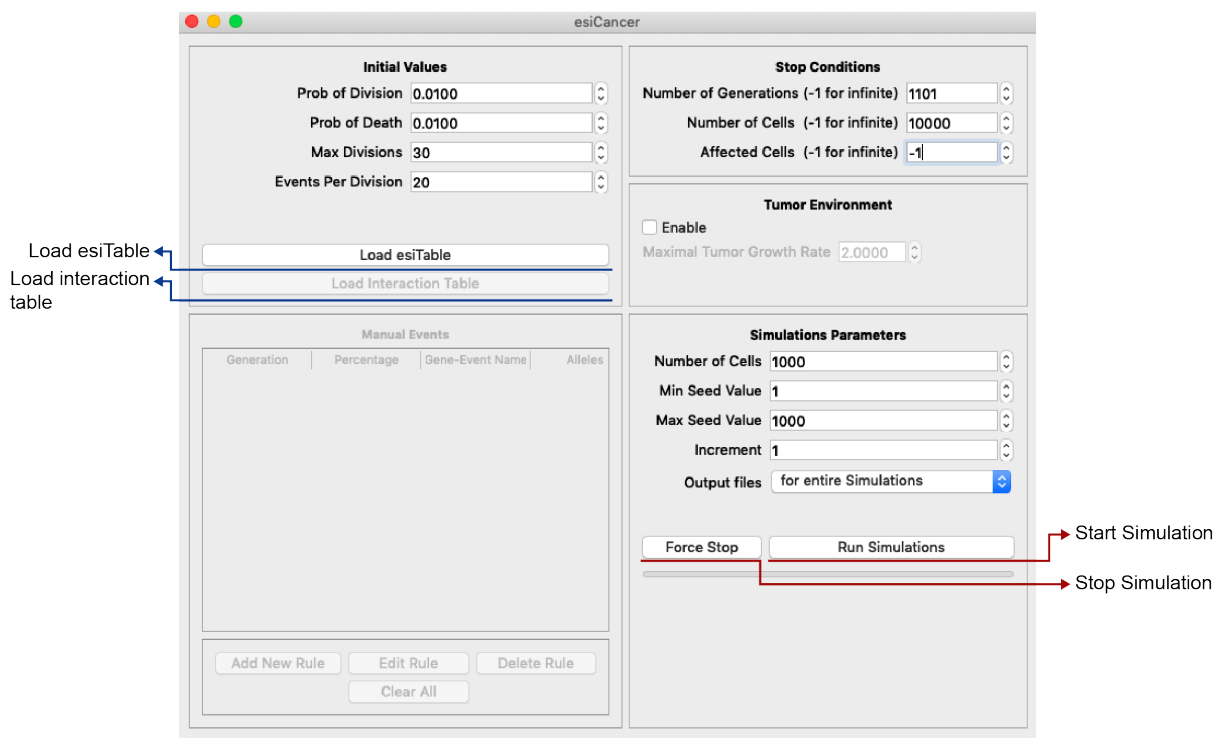
Initial Values

- **Prob. of Division:** initial probability of division for all cells from initial population

- **Prob. of Death:** initial probability of death for all cells from initial population

- **Max Divisions:** initial maximal number of divisions that a cell can go through

- **Events per Division:** initial number of events to be inserted into each cell at each division

Stop Conditions

- **Number of Generations:** number of maximum generations allowed per run

- **Number of Cells:** maximum number of cells that a simulation can reach

- **Affected Cells:** maximum number of affected cells that a simulation can reach

Tumor Environment

- **Maximal Tumor Growth Rate:** maximal population growth rate allowed per generation

$$MTGR = \frac{population\ Gen_n}{population\ Gen_{n-1}}$$

When enable, if MTGR is above the maximal allowed, affected cells are eliminated randomly to reduce MTGR.

$$Number\ Eliminated\ Cells = population\ Gen_n - (population\ Gen_{n-1} * MTGR)$$

Simulations parameters

- **Number of Cells:** starting number of cells

- **Min Seed Value:** starting seed value

- **Max Seed Value:** ending seed value

- **Increment:** defines the increment to be used from minimum to maximum

- **Output files:** defines the format of output files

  - **for each Seed:** individual output file for each seed simulated

  - **for entire Simulation:** output file for all seeds simulated in a single summarized format

  - **both:** create both types of output for all seeds simulated

OBS: for single seed simulation, set min and max seed value as the same

3.2 Manual Events Parameters
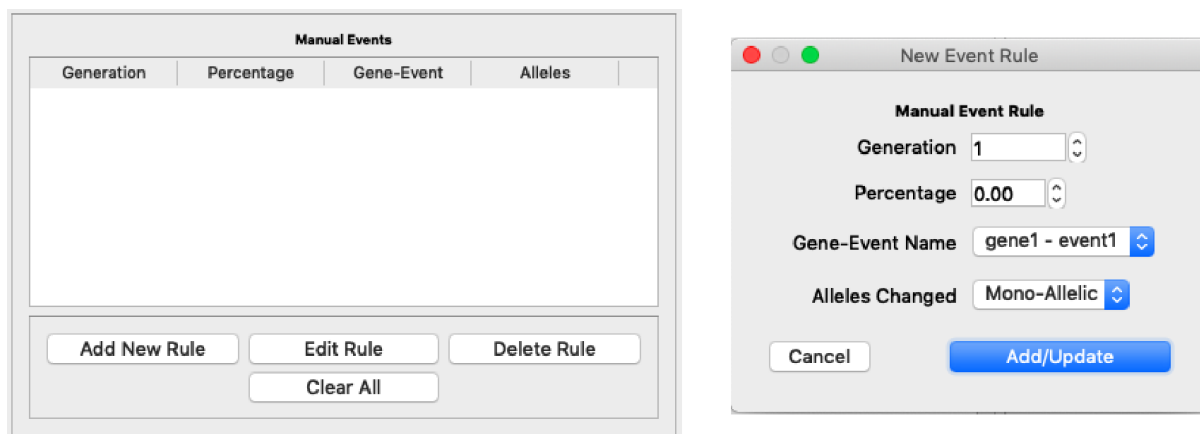
Manual Events

- **Generation:** generation that event rule is added

- **Percentage:** percentage of cells that event rule is added

- **Gene-Event Name:** Gene-event that is added event rule (according to esiTable)

- **Alleles Changed:** Number of alleles affected by event rule(mono or bi-allelic)

4 Tables

4.1 esiTable Parameters

genome size	GENES	EVENTS	PROBEVENT	DOMINANCE	DIVFUNC	DIVRATE	DEFUNC	DEATHRATE	MUTFUNC	MUT	MAXDIVFUNC	MAXDIVRATE	MEFUNC	MICROENVIRONMENT
1000	gene1	event1	1	0	MULT	1.5	MULT	1	MULT	1	ADD	0	MULT	1
	gene2	event1	1	0.3	MULT	1.5	MULT	1	MULT	1	ADD	0	MULT	1
	gene3	event1	1	0.3	MULT	1.5	MULT	1	MULT	1	ADD	0	MULT	1
	gene3	event2	1	0.3	MULT	1.5	MULT	1	MULT	1	ADD	0	MULT	1
	gene4	event1	1	1	MULT	1.5	MULT	1	MULT	1	ADD	0	MULT	1
	gene5	event1	2	1	MULT	1.1	MULT	1	MULT	1	ADD	0	MULT	1
	gene6	event1	1	1	MULT	1.1	MULT	1	MULT	1	ADD	0	MULT	1
	gene7	event1	0	1	MULT	1.1	MULT	1	MULT	1	ADD	0	MULT	1
	gene8	event1	1	1	MULT	1.1	MULT	1	MULT	1	ADD	0	MULT	1
	deletion	event	1	1	MULT	1	MULT	1	MULT	1	ADD	0	MULT	1

genome size - sum of the oncogenome and the normal genome

Genes - name of a given gene

Events - name of a given event

ProbEvent - probability of a specific event to occur (Prob = ProbEvent / genome size)

Dominance - the impact of a specific event if it occurs in one allele

DivFunc - determines if the event adds or multiplies a specific value to cell's division rate

DivRate - the impact of a specific event on cell's division rate

Defunc - determines if the event adds or multiplies a specific value to cell's death rate

DeathRate - the impact of a specific event on cell's death rate

MutFunc - determines if the event adds or multiplies a specific value to cell's events per division rate

MaxDivFunc - determines if the event adds or multiplies a specific value to cell's maximum number of division rate

MaxDivRate - the impact of a specific event on cell's maximum number of division rate

MeFunc - determines if the event adds or multiplies a specific value to MTGR

MicroEnviroment - the impact of a specific event on MTGR

4.2 Interaction Table Parameters

Gene_Before	Event_Before	Gene_After	Event_After	DIVMOD	DEMOD	MAXDIVMOD	Link
gene1	event1	gene2	event1	0	0	0	0
gene2	event1	gene3	event1	0	0	0	0
gene1	event1	gene5	event1	5	1	1	0
gene3	event2	gene5	event1	3	1	1	0
gene3	event1	gene4	event1	10	1	1	0
deletion	event	gene6	event1	1	1	1	1
deletion	event	gene7	event1	1	1	1	1
deletion	event	gene8	event1	1	1	1	1

Gene_Before - the gene name affected first

Event_Before - the event name affected first

Gene_After - the gene name affected after the first

Event_After - the event name affected after the first

DIVMOD - the impact of interaction on cell's division rate

DEMOD - the impact of interaction on cell's death rate

MAXDIVMOD - the impact of interaction on cell's maximum number of division rate

Link - determines if an event affects more than one gene (**1 = affect; 0 = does not affect**)

5   Output

5.1   for each Seed

5.1.1   _seed_ancestralResults.csv

Generation	Population	A0	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	A15	A16	A17	A18	A19	A20
0	1000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	1008	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1
2	1007	1	1	1	1	1	1	0	1	1	0	1	1	1	1	1	0	1	1	1	1	1
3	1007	1	1	1	1	0	1	0	1	1	0	1	1	1	1	1	0	0	1	1	1	1
4	1004	1	1	1	1	0	1	0	1	1	0	1	1	1	1	1	0	0	1	1	1	1
5	1010	1	1	1	1	0	1	0	1	1	0	1	1	1	1	1	0	0	1	1	1	1
6	1018	1	1	1	1	0	1	0	1	1	0	1	1	2	1	1	0	0	1	1	1	1
7	1017	1	1	0	1	0	1	0	1	1	0	1	1	2	1	1	0	0	1	1	1	1
8	1022	1	1	0	1	0	1	0	1	1	0	1	1	2	1	1	0	0	1	1	1	1
9	1019	1	1	0	1	0	1	0	1	1	0	1	1	2	1	1	0	0	1	1	1	1
10	1020	1	1	0	1	0	1	0	1	1	0	1	1	2	1	1	0	0	1	1	1	1
11	1020	1	1	0	1	0	1	0	1	1	0	1	1	2	1	1	0	0	1	1	1	1
12	1018	1	1	0	1	0	1	0	1	1	0	1	1	2	1	1	0	0	1	1	1	1
13	1024	1	1	0	1	0	1	0	1	2	0	1	1	2	1	1	0	0	1	1	1	1
14	1024	1	1	0	1	0	1	0	1	2	0	1	1	2	1	1	0	0	1	1	1	1
15	1023	1	1	0	1	0	1	0	1	2	0	1	1	2	1	1	0	0	1	1	1	1

For each generation returns the **total population** and the **number of descendants** per cell from initial population (column A0-An).

5.1.2   _seed_eventsMutationResults.csv

Generation	gene1 - event1	gene2 - event1	gene3 - event1	gene3 - event2	gene4 - event1	gene5 - event1	gene6 - event1	Total Population
0	0	0	0	0	0	0	0	1000
1	0	0	0	0	0	0	1	1008
2	0	0	0	0	1	0	1	1007
3	0	1	1	0	1	0	1	1007
4	1	1	2	0	1	0	1	1004
5	1	3	2	0	1	0	1	1010
6	1	3	2	0	1	1	1	1018
7	1	4	2	1	1	1	2	1017
8	2	4	2	1	1	2	2	1022
9	3	4	2	1	1	2	2	1019
10	3	4	2	1	1	2	2	1020
11	3	4	2	1	1	2	2	1020
12	3	4	2	1	1	2	2	1018
13	3	4	2	1	3	2	2	1024
14	2	4	2	1	3	2	2	1024
15	2	4	2	1	4	3	2	1023

For each generation returns the **total population** and the **number of cells affected in each event**. In this case, cells with event affected in one or two alleles will be accounted equally.

5.1.3 _seed_genesMutationResults.csv

Generation	gene1 monoAl	gene1 biAl	gene2 monoAl	gene2 biAl	gene3 monoAl	gene3 biAl	gene4 monoAl	gene4 biAl	gene5 monoAl	gene5 biAl	gene6 monoAl	gene6 biAl	Total Population
0	0	0	0	0	0	0	0	0	0	0	0	0	1000
1	0	0	0	0	0	0	0	0	0	0	1	0	1008
2	0	0	0	0	0	0	1	0	0	0	1	0	1007
3	0	0	1	0	1	0	1	0	0	0	1	0	1007
4	1	0	1	0	2	0	1	0	0	0	1	0	1004
5	1	0	3	0	2	0	1	0	0	0	1	0	1010
6	1	0	3	0	2	0	1	0	1	0	1	0	1018
7	1	0	4	0	3	0	1	0	1	0	2	0	1017
8	2	0	4	0	3	0	1	0	2	0	2	0	1022
9	3	0	4	0	3	0	1	0	2	0	2	0	1019
10	3	0	4	0	3	0	1	0	2	0	2	0	1020
11	3	0	4	0	3	0	1	0	2	0	2	0	1020
12	3	0	4	0	3	0	1	0	2	0	2	0	1018
13	3	0	4	0	3	0	3	0	2	0	2	0	1024
14	2	0	4	0	3	0	3	0	2	0	2	0	1024
15	2	0	4	0	3	0	4	0	3	0	2	0	1023

For each generation returns the **total population** and the **number of cells affected in each gene** in one or two alleles. In this case, cells affected in only one allele will be accounted in monoAl column and cells affected in both alleles in biAl.

5.1.4 _parameters.txt

```
Parameters
Seed: 3
Number of cells (started): 1000
StdProbOfProliferate: 0.01
StdProbOfDeath: 0.01
Telomeres: 30
Events per div: 20
```

Returns the input parameters (seed, Number of Cells, Proliferation Rate, Death Rate, Maximum Number of Division, Events per Division). It is used as a control for simulations.

5.2 for entire Simulation

5.2.1 _automatics_ancestralResults.csv

Seed	Generation	NumberOfDivisions	Population	A0	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	A15	A16	A17	A18	A19	A20
1	461	17678	10174	0	0	0	0	0	0	29	12	0	71	0	0	0	0	18	0	15	0	0	0	0
2	424	17864	10149	0	17	52	3	0	0	0	0	1	0	0	0	0	0	0	0	0	1	9	20	0
3	417	17636	10201	0	0	2	0	0	0	0	0	4	0	0	0	16	0	0	0	22	0	0	2	0
4	489	19151	10078	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	45	0
5	509	19988	10023	0	0	0	0	0	0	0	0	0	0	0	114	0	0	0	0	0	0	0	0	0
6	455	17380	10152	0	0	0	0	0	0	0	0	15	0	0	0	0	0	0	0	5	0	0	4	0
7	471	18085	10029	0	0	0	0	0	0	0	0	0	12	0	0	0	0	8	0	0	0	0	0	1
8	499	19885	10179	159	0	0	0	0	0	100	0	0	6	12	9	0	0	0	0	0	0	0	35	0
9	461	18058	10113	0	0	96	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	1	0	0
10	514	19727	10019	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	21	0	0	0
11	499	19294	10045	0	0	0	0	0	0	0	0	0	0	0	19	0	0	0	0	0	0	8	0	0
12	423	17205	10196	0	0	0	0	0	0	0	0	18	13	0	0	1	0	0	0	15	0	0	0	0
13	466	19087	10060	0	0	0	10	40	0	0	3	382	4	0	0	0	0	0	0	0	0	0	0	10
14	504	19376	10071	12	0	194	0	0	0	47	0	0	0	0	0	0	0	0	0	3	0	0	0	0
15	524	20259	10029	0	0	0	0	4	0	0	0	0	76	0	0	0	0	5	0	4	0	0	0	0
16	486	18915	10032	236	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	11	3	0	0	6
17	398	17145	10031	0	4	0	6	0	0	0	0	0	0	0	0	7	4	0	0	0	0	2	1	0
18	393	15718	10138	2	5	0	5	0	0	0	0	0	4	0	0	0	24	0	0	4	6	0	0	0
19	502	19702	10065	0	0	0	15	0	0	197	0	0	376	0	16	0	2581	0	0	0	3	0	0	8
20	432	18501	10002	0	2	0	0	0	0	0	0	0	0	1	0	0	5	0	0	0	0	0	0	0

For each seed returns the **total population**, the **total number of divisions** and the **final generation**, which correspond to generation that reach one of the stop conditions. Further, for each cell from initial population, returns the **number of descendants** on final population.

5.2.2 _automatics_eventsMutationResults.csv

Seed	Generation	gene1 - event1	gene2 - event1	gene3 - event1	gene3 - event2	gene4 - event1	gene5 - event1	gene6 - event1	Total Population
1	461	2954	7838	7423	7570	8130	8201	4114	10174
2	424	3219	6818	7629	3559	8990	7711	5183	10149
3	417	5312	7437	6572	6680	8562	8787	4282	10201
4	489	3913	5635	5196	4881	8767	6178	4973	10078
5	509	3284	4464	4377	5843	8327	7061	4161	10023
6	455	6707	8158	7270	6407	8941	5412	4931	10152
7	471	6337	4789	3950	6576	8231	7226	3042	10029
8	499	2792	6180	4902	6441	7760	7107	5381	10179
9	461	3345	7806	3588	6979	7748	7902	7370	10113
10	514	3898	5402	6604	5412	7924	6100	6553	10019
11	499	5786	6279	3262	6453	7872	6553	4463	10045
12	423	3806	7151	7398	5118	8948	7075	5919	10196
13	466	4281	5572	6430	6146	7626	6326	4543	10060
14	504	3476	6958	3071	6551	8463	7368	3948	10071
15	524	3094	5358	5608	4284	8043	6249	4220	10029
16	486	4730	6149	6531	3670	8241	7029	4704	10032
17	398	2439	5995	8389	7667	8532	8314	2421	10031
18	393	8377	7972	6717	1377	8883	4479	8283	10138
19	502	4966	5575	3759	6759	8117	7516	4502	10065
20	432	3137	4668	6144	7161	8385	5616	6150	10002

For each seed returns the **total population**, the **final generation** and the **number of cells affected in each event** on final population. In this case, cells with event affected in one or two alleles will be accounted equally.

5.2.3 _automatics_genesMutationResults.csv

Seed	Generation	gene1 monoAl	gene1 biAl	gene2 monoAl	gene2 biAl	gene3 monoAl	gene3 biAl	gene4 monoAl	gene4 biAl	gene5 monoAl	gene5 biAl	gene6 monoAl	gene6 biAl	Total Population
1	461	2341	613	1568	6270	1558	7171	6394	1736	6480	1721	3640	474	10174
2	424	2340	879	2157	4661	2846	5760	5163	3827	5906	1805	4257	926	10149
3	417	1775	3537	5884	1553	1940	6409	7194	1368	4981	3806	3432	850	10201
4	489	2339	1574	3830	1805	3107	4587	6752	2015	4295	1883	4169	804	10078
5	509	2623	661	2878	1586	3135	4864	6336	1991	5335	1726	3691	470	10023
6	455	3881	2826	2571	5587	2394	5847	7320	1621	4099	1313	4488	443	10152
7	471	1012	5325	3654	1135	1810	6643	6908	1323	5080	2146	2785	257	10029
8	499	2490	302	3791	2389	2173	6235	6790	970	5684	1423	3010	2371	10179
9	461	2650	695	931	6875	2383	5885	6763	985	4947	2955	6024	1346	10113
10	514	3136	762	4047	1355	1690	6503	7066	858	4943	1157	6031	522	10019
11	499	3218	2568	4314	1965	2520	5676	5789	2083	5105	1448	3827	636	10045
12	423	2254	1552	913	6238	2808	5306	6481	2467	5618	1457	5266	653	10196
13	466	2074	2207	4379	1193	2011	6319	6616	1010	5025	1301	4155	388	10060
14	504	2804	672	4509	2449	2774	5247	4247	4216	5151	2217	3397	551	10071
15	524	1934	1160	3995	1363	3228	4249	6904	1139	4632	1617	3886	334	10029
16	486	2855	1875	3878	2271	2415	5482	7418	823	4249	2780	4360	344	10032
17	398	2187	252	4383	1612	870	7933	7005	1527	3966	4348	2157	264	10031
18	393	677	7700	634	7338	1680	6128	7548	1335	3565	914	7204	1079	10138
19	502	2566	2400	4082	1493	3571	4551	6773	1344	2999	4517	3871	631	10065
20	432	2553	584	2858	1810	2786	5611	3953	4432	4406	1210	5691	459	10002

For each seed returns the **total population**, the **final generation** and the **number of cells affected in each gene** in one or two alleles on final population. In this case, cells affected in only one allele will be accounted in monoAl column and cells affected in both alleles in biAl.

5.2.4 _automatics_sequenceEachCell.csv

Seed	Generation	Population	Cell0 firstStrand	Cell0 secondStrand	Cell1 firstStrand	Cell1 secondStrand	Cell2 firstStrand	Cell2 secondStrand	Cell3 firstStrand	Cell3 secondStrand
1	462	10174	1s000000000	2s000000000	1s011000000	2s011100000	1s000111000	2s001000000	1s000000000	2s000000000
2	425	10149	1s011111000	2s011100000	1s000010000	2s001000000	1s100111000	2s010110010	1s011110000	2s010000000
3	418	10201	1s001000000	2s010010000	1s000010000	2s000000000	1s100100011	2s100011001	1s000000000	2s000000000
4	490	10078	1s000100000	2s001000000	1s000100000	2s011000001	1s001101010	2s111101011	1s000110011	2s000001001
5	510	10023	1s000000000	2s000100000	1s100001000	2s001000000	1s000000000	2s000000001	1s000000000	2s000100000
6	456	10152	1s000000000	2s000000000	1s000001000	2s010100000	1s000000010	2s010101000	1s000000010	2s000000000
7	472	10029	1s001010000	2s100010010	1s001001000	2s001101000	1s010000001	2s001000011	1s000010000	2s000010000
8	500	10179	1s001000000	2s000000000	1s000010011	2s001000000	1s001010000	2s001100010	1s110100001	2s001000000
9	462	10113	1s001010000	2s000000010	1s000000000	2s000010010	1s011100000	2s010000001	1s001010000	2s011000010
10	515	10019	1s000010000	2s000010001	1s100100001	2s000000000	1s111001000	2s001110011	1s000000000	2s000001000
11	500	10045	1s011000000	2s111000000	1s101111001	2s111010000	1s011000000	2s000010010	1s001101011	2s001100011
12	424	10196	1s000000000	2s000001000	1s001110010	2s001100011	1s000000000	2s000001000	1s011011000	2s011100010
13	467	10060	1s010010000	2s000000000	1s101100000	2s101010001	1s000010000	2s000000000	1s110100010	2s000100001
14	505	10071	1s000000000	2s000000000	1s000010000	2s001100000	1s010000010	2s000100000	1s000010000	2s111111000
15	525	10029	1s000000010	2s000010000	1s000010001	2s001110000	1s000000000	2s000000000	1s100110010	2s010001000
16	487	10032	1s001110000	2s000001001	1s000110000	2s000001001	1s111110010	2s100011000	1s011110010	2s011011001
17	399	10031	1s000010000	2s100000000	1s000000000	2s000100000	1s000110000	2s001000000	1s000000001	2s011010010
18	394	10138	1s110101010	2s110100001	1s110111000	2s111000001	1s010000000	2s000000000	1s000000000	2s000000000
19	503	10065	1s101010010	2s001110000	1s100010000	2s001100000	1s011010010	2s001110000	1s101110000	2s111010011
20	433	10002	1s000000000	2s000100000	1s010010000	2s100100000	1s001100000	2s000111000	1s001100000	2s101101000

For each seed returns the **total population**,the **final generation** and the **single cell genes sequence** as two independent list in a binary form (**1 = affected; 0 = not affected**) for each cell from final population. The first strand is signalized as 1s and the second strand as 2s followed by the binary sequence.

6 Support and Bugs

If you are having issues, please contact Bernardo Henz bernardohenz@gmail.com

7 License

esiCancer is under GNU Public License v3.0.