Developing Data-Driven Models for Obtaining Personalized Cancer Treatments

Leili Shahriyari
Department of Mathematics and Statistics
University of Massachusetts Amherst

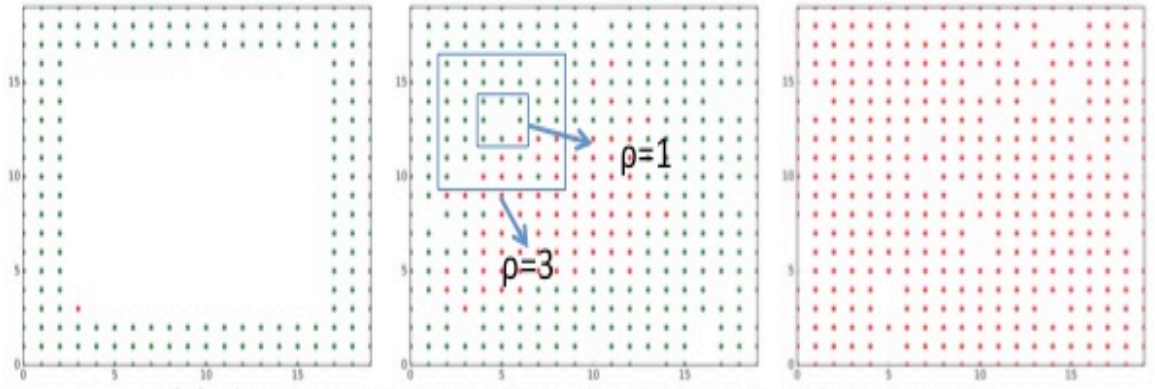
MPIs Souvik Roy and Suvra Pal
Department of Mathematics, University of Texas Arlington

ITCR teleconference
Nov 1, 2019

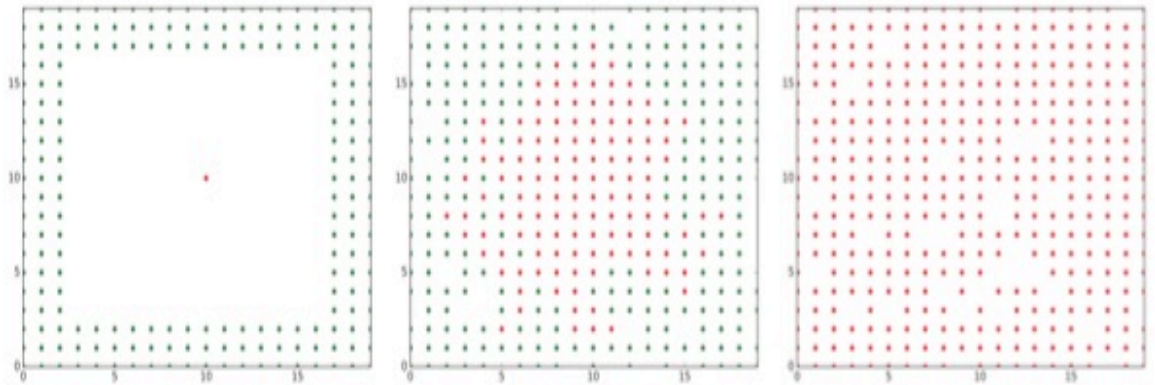
Overview

- Background – Wound healing process during and after treatments
- Algorithm
- FOLFIRI treatment
- Mathematical model
- Tumor deconvolution
- Data normalization
- Parameter estimation

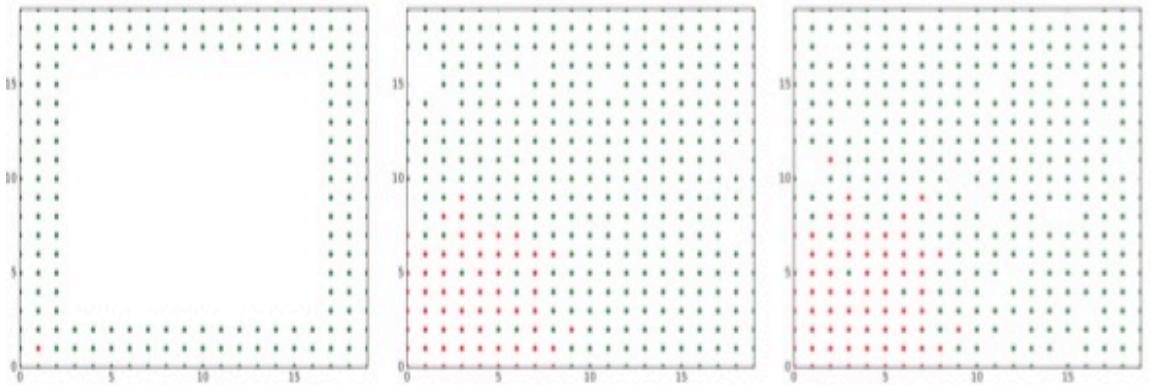
Spatial Model Wound healing process after treatments.



(a) A single cancer cell in the boundary of the wound



(b) A single cancer cell in the middle of the wound

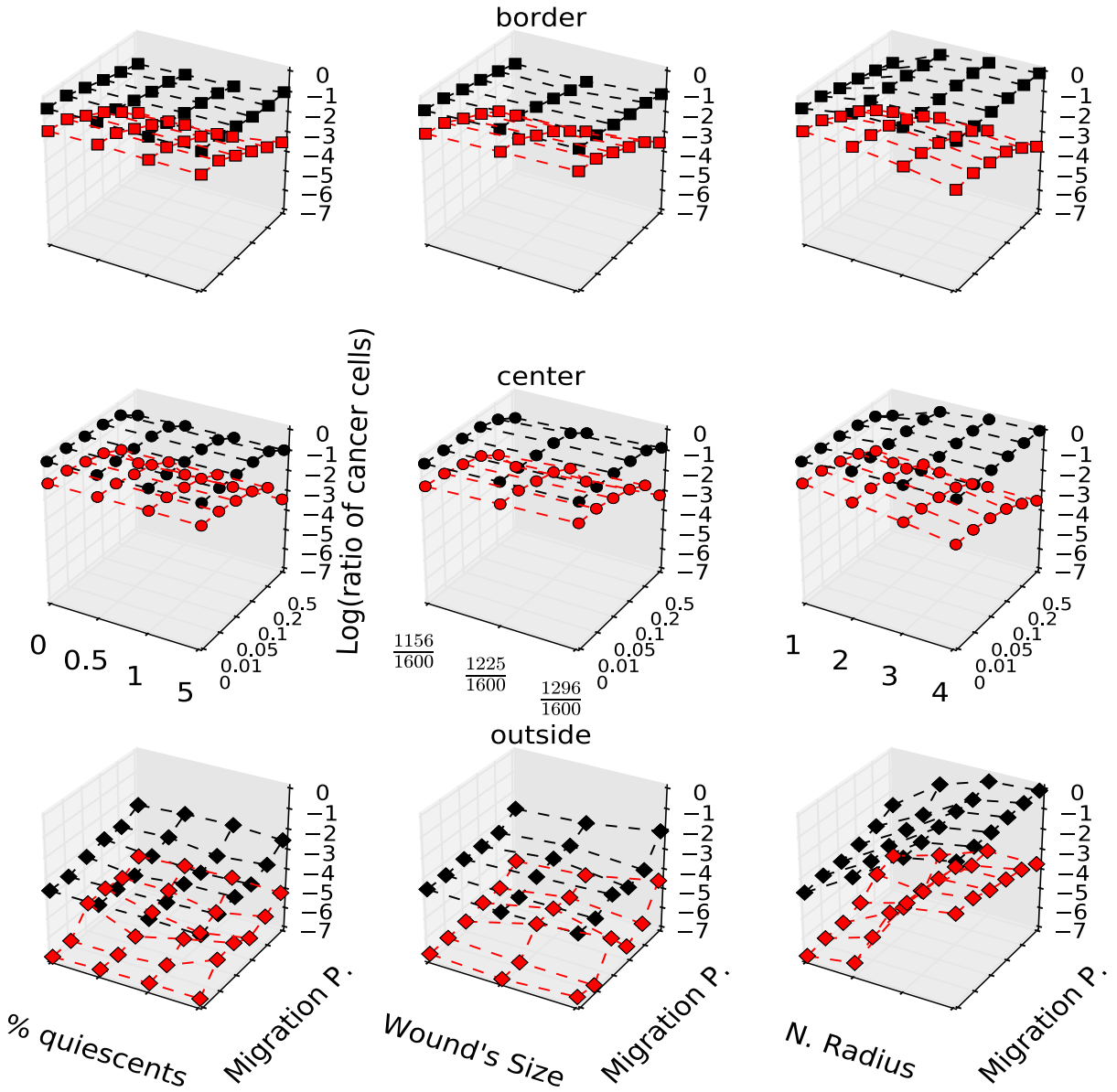


(c) A single cancer cell outside of the wound's boundary

Shahriyari L. (2017) Royal Society Interface

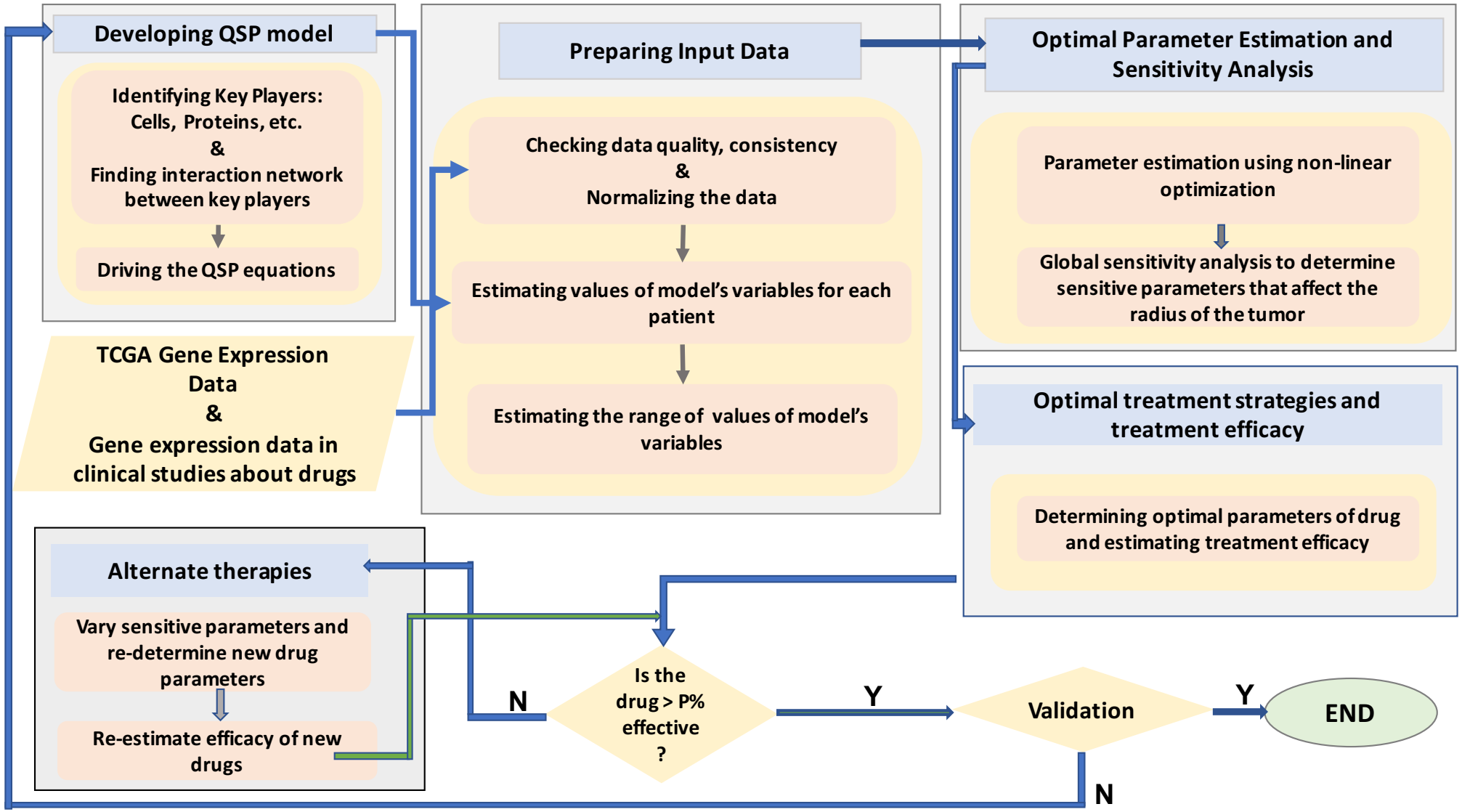
Ratio of cancer cells after treatments

Red: after the wound has been healed
Black: The final time of simulations T, where $T = 10D$ number of updating time steps, and D is the tissue's size.



Shahriyari L. (2017) Royal Society Interface

Algorithm


Developing QSP model
Identifying Key Players: Cells, Proteins, etc. & Finding interaction network between key players
Driving the QSP equations
TCGA Gene Expression Data & Gene expression data in clinical studies about drugs
Preparing Input Data
Checking data quality, consistency & Normalizing the data
Estimating values of model's variables for each patient
Estimating the range of values of model's variables
Optimal Parameter Estimation and Sensitivity Analysis
Parameter estimation using non-linear optimization
Global sensitivity analysis to determine sensitive parameters that affect the radius of the tumor
Optimal treatment strategies and treatment efficacy
Determining optimal parameters of drug and estimating treatment efficacy
Alternate therapies
Vary sensitive parameters and re-determine new drug parameters
Re-estimate efficacy of new drugs
Is the drug > P% effective ?
N
Y
Validation
Y
END
N

FOLFIRI treatment

FOLFIRI is made up of:

- folinic acid (also called leucovorin, calcium folinate or FA)
- fluorouracil (also called 5FU)
- Irinotecan

Patients have FOLFIRI chemotherapy as cycles of treatment. Each cycle of treatment lasts 2 weeks (14 days). Depending on their needs, they may have up to 12 cycles, taking up to 6 months in total.

FOLFIRI treatment

Day 1

- Patients have irinotecan through a drip into the bloodstream over 60 to 90 minutes
- Patients have folinic acid through a drip into the bloodstream over 2 hours
- Patients have an injection of fluorouracil into the bloodstream over 5 minutes
- Patients have an infusion of 5FU through a drip or pump into the bloodstream for 46 hours

Day 2

- Patients continue to have the 5FU infusion

Day 3 to 14

- no treatment

https://www.cancerresearchuk.org/

QSP Model - example

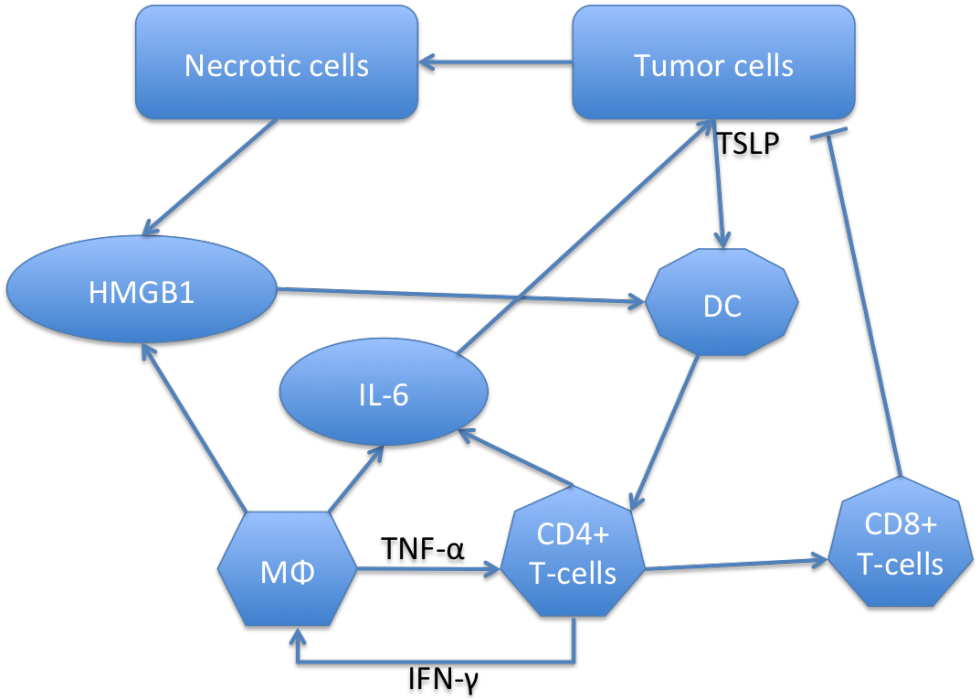
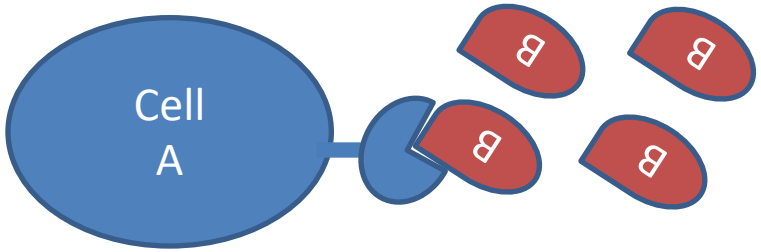


Table 1: Model's Variables in units of g/cm^3.

Symbol	Definition
N	Density of necrotic cells (Cells going through necrosis)
C	Density of tumor cells
D	Density of dendritic cells
T_4	Density of $CD4^+$ T-cells (Th1, Th2, and Th17)
T_8	Density of $CD8^+$ T-cells
M	Density of M_1 macrophages
H	Concentration of HMGB1
I_6	Concentration of IL-6

Michaelis-Menten law

- An A-cell becomes activated when a molecule of B ligands to a cytoplasmic-receptor on the A-cell, thereby resulting in an activated A-cell, which we designate as C-cell.



Instead of the mass action law, we use the Michaelis-Menten law.

- $dC/dt = \lambda B A / (A+K)$
- where K and λ are constant. The constant K is called the half-saturation of A, and we assume that it is equal to the steady state of A.

Modeling tumor

We extend the ODE system to a system of PDEs to include the spatial proliferation of cancer cells. We assume that the tumor occupies a region $\Omega(t)$ and that the combined density of cells at each point in $\Omega(t)$ is approximately equal to 1 g/cm^3, so that

$$\frac{D + T_4 + M + T_8 + N + C}{|\Omega(t)|} = 1 \ g/cm^3. \tag{1}$$

Modeling tumor

Since cancer growth is abnormal, the assumption (1) implies that a velocity field $\mathbf{v}$, by which the cells are moving within $\Omega(t)$, is developed. Hence, for each species X of cells, the equation

$$\frac{dX}{dt} = F$$

from the ODE model is replaced by the equation

$$\frac{\partial X}{\partial t} + div(\mathbf{v}X) = D_X \nabla^2 X + F$$

where D_X is the dispersion (or diffusion) coefficient of species X.

The cytokines H and I_6 are also diffusing with their own diffusion coefficients. Since these coefficients are much larger than cells' diffusion coefficients, we may neglect the effect of $\mathbf{v}$, and thus replace each equation $\frac{dX}{dt} = F$ by $\frac{\partial X}{\partial t} = D_X \nabla^2 X + F$.

$$\frac{\partial D}{\partial t} + div(\mathbf{v}D) = D_D\nabla^2 D + \overbrace{A_D}^{\text{Rate of constitutively produced dendritic cells}} + \underbrace{D^0(\lambda_{DH}\frac{H}{K_H+H} + \lambda_{DC}\frac{C}{K_C+C})}_{\text{Activation of DCs by HMBG1 and tumor}} - \underbrace{\delta_D D}_{\text{Death rate}} \tag{2}$$

$$\frac{\partial T_4}{\partial t} + div(\mathbf{v}T_4) = D_{T_4}\nabla^2 T_4 + \underbrace{\lambda_{T_4 D}T_4^0\frac{D}{K_D+D}}_{\text{Activation by DCs}} + \underbrace{\lambda_{T_4 M}T_4\frac{M}{K_M+M}}_{\text{Proliferation by macrophages}} - \underbrace{\delta_{T_4}T_4}_{\text{Death rate}} \tag{3}$$

$$\frac{\partial M}{\partial t} + div(\mathbf{v}M) = D_M\nabla^2 M + \overbrace{A_M}^{\text{Rate of constitutively produced macrophages}} + \underbrace{\lambda_{MT_4}M\frac{T_4}{K_{T_4}+T_4}}_{\text{Activation by CD4}^+\text{ T-cells}} - \underbrace{\delta_M M}_{\text{Death rate}} \tag{4}$$

$$\frac{\partial T_8}{\partial t} + div(\mathbf{v}T_8) = D_{T_8}\nabla^2 T_8 + \overbrace{\lambda_{T_8T_4}T_8^0\frac{T_4}{K_{T_4}+T_4}}^{\text{Activation by CD4}^+\text{ T-cells}} - \underbrace{\delta_{T_8}T_8}_{\text{Death rate}} \tag{5}$$

$$\frac{\partial N}{\partial t} + div(\mathbf{v}N) = D_N\nabla^2 N + \underbrace{\frac{1}{4}\delta_{CT_8}T_8C}_{\text{Production by CD8}^+\text{ T-cells}} + \overbrace{\delta_C C}^{\text{Production by tumor}} - \underbrace{\delta_N N}_{\text{Death rate}} \tag{6}$$

$$\frac{\partial H}{\partial t} = D_H\nabla^2 H + \overbrace{\lambda_{HN}N}^{\text{Production by Necrotic cells}} + \underbrace{\lambda_{HM}M}_{\text{Production by M}\Phi} - \underbrace{\hat{\lambda}_{DH}D^0\frac{H}{K_H+H}}_{\text{Absorption by DCs}} - \underbrace{\delta_H H}_{\text{Degradation}} \tag{7}$$

$$\frac{\partial I_6}{\partial t} = D_{I_6}\nabla^2 I_6 + \overbrace{\lambda_{I_6T_4}T_4}^{\text{Production by } T_4\text{ cells}} + \underbrace{\lambda_{I_6M}M}_{\text{Production by M}\Phi} - \underbrace{\hat{\lambda}_{CI_6}C\frac{I_6}{K_{I_6}+I_6}}_{\text{Prolif. of cancer cells by IL-6}} - \underbrace{\delta_{I_6}I_6}_{\text{Degradation}} \tag{8}$$

$$\frac{\partial C}{\partial t} + div(\mathbf{v}C) = D_C\nabla^2 C + \underbrace{\lambda_C C(1-\frac{C}{C_0})}_{\text{Proliferation}} - \overbrace{\delta_{CT_8}T_8C}^{\text{Death by CD8}^+\text{ T-cells}} + \underbrace{\lambda_{CI_6}C\frac{I_6}{K_{I_6}+I_6}}_{\text{Prolif. of cancer cells by IL-6}} - \overbrace{\delta_C C}^{\text{Death rate}} \tag{9}$$

Modeling tumor

Boundary Conditions

We assume that T-cells migrate from the lymph nodes into the cancer, and express this by the boundary conditions

$$\frac{\partial D}{\partial n} + \alpha(D - D^0) = 0 \ , \ \frac{\partial T_4}{\partial n} + \alpha(T_4 - T_4^0) = 0 \ , \ \frac{\partial T_8}{\partial n} + \alpha(T_8 - T_8^0) = 0$$

where D^0, T_4^0 and T_8^0 are the half-saturation values of D, T_4 and T_8. All other cells satisfy the no-flux condition:

$$\frac{\partial M}{\partial n} = \frac{\partial N}{\partial n} = \frac{\partial C}{\partial n} = 0.$$

We also assume no-flux for the cytokines:

$$\frac{\partial H}{\partial n} = \frac{\partial I_6}{\partial n} = 0.$$

Modeling tumor

We assume that all cells have approximately the same volume and surface area, so that their dispersion coefficients are the same. By adding the equations of cells, we get an equation for $\nabla \cdot \mathbf{v}$:

$$\nabla \cdot \mathbf{v} = \sum \Big[\text{RHS of Eqs. (2)-(6) and (9)}\Big].$$

We also assume that the tumor's region $\Omega(t)$ is a sphere $0 \leq r \leq R(t)$, and all the variables are spherically symmetric. Additionally, we assume that the free boundary $r = R(t)$ moves with the velocity of the cells, that is

$$\frac{dR}{dt} = \mathbf{v} \cdot \mathbf{e}_r = v(R(t), t),$$

where $\mathbf{e}_r = \frac{\mathbf{x}}{|\mathbf{x}|}$ is the radial unit vector.

Modeling treatment

- Finding optimal treatment

$$\frac{\partial \boldsymbol{A}(\boldsymbol{\theta})}{\partial t} = D_{\boldsymbol{A}} \Delta \boldsymbol{A} + F(\boldsymbol{A}, \boldsymbol{\theta}) + \eta(t).$$

- Cross Validation.

Estimating values of variables at one time point

Analytical tools developed to provide an estimation of the abundances of member cell types in a mixed cell population, using gene expression data, including:

- DeconRNASeq (Gong et al. 2013)
- CIBERSORT (Newman et al. 2015)
- Method by (Senbabaoglu et al. 2016) using single-sample GSEA (ssGSEA) score (Barbie et al. 2009).
- The same method using singscore (Foroutan et al. 2018) instead of ssGSEA score.

- **DREAM Challenge**
- Generating test cases (generating artificial data points using a signature matrix)

Normalizing Gene Expression Data

- Raw data $D=[g_{i,j}]=[P_i]=[G_j]$, $P_i=[g_{i,j} \quad 1\leq j \leq 60483]$

Excluded genes with zero values in the entire D

$P_i=[g_{i,j} \quad 1\leq j \leq 57813]$, $G_j=[g_{i,j} \quad 1\leq i \leq n]$,

n= number of files

Normalization methods:

- *Scaling:*

$P_i=[g_{i,j}/\max(P_i) \; 1\leq j \leq 57813]$ or $G_j=[g_{i,j}/\max(G_j) \quad 1\leq i \leq n]$

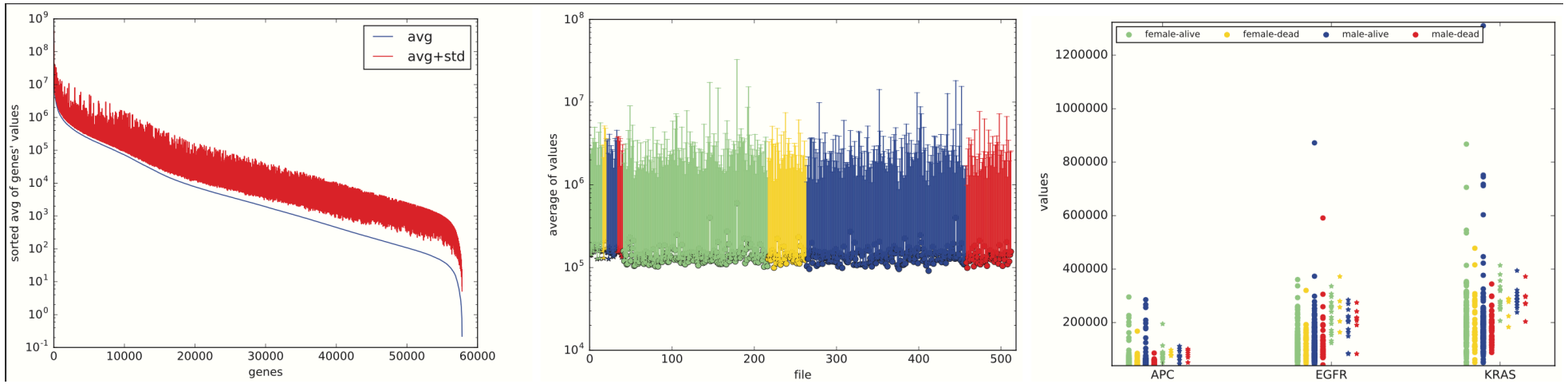
- *Unit length:*

$P_i=[g_{i,j}/|P_i| \quad 1\leq j \leq 57813]$ or $G_j=[g_{i,j}/|G_j| \quad 1\leq i \leq n]$

- *Z-score*: $P_i=$ z-score(P_i) or $G_j=$ z-score(G_j)

Statistical Analysis

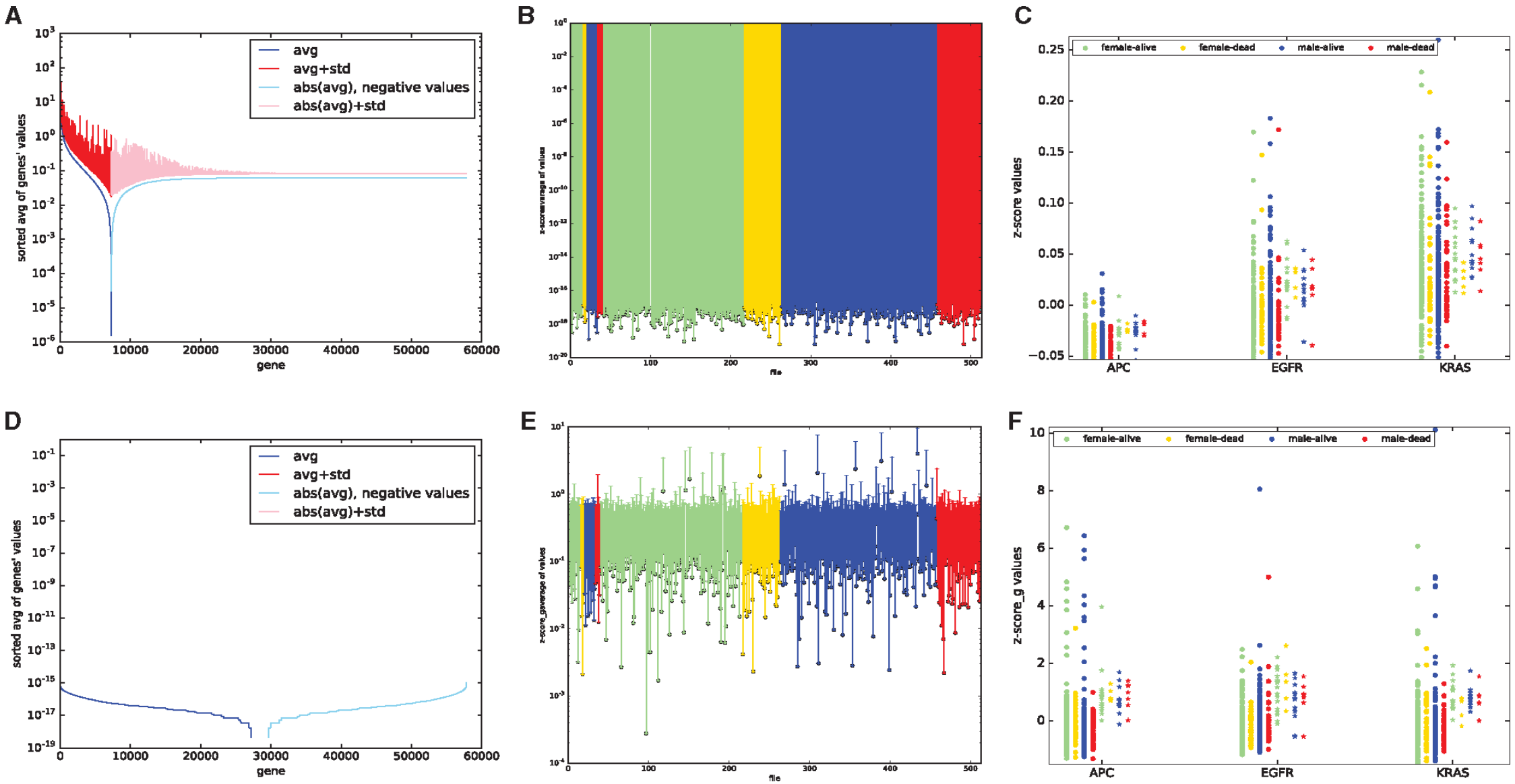
Normalization method	avg(avg(g_j))=avg(avg(P_i))	avg(std(g_j))	avg(std(P_i))	cor(EGFR, KRAS)
Raw data	142069.91	136066.44	2784854.65	0.226
Scaling files	0.000599	0.000522	0.0092737	0.51
Scaling genes	0.075386	0.084623	0.131425	0.226
Scaling files to the unit length	0.000258	0.000205	0.004150	0.43
Scaling genes to the unit length	0.017967	0.036806	0.03099	0.226
Standardizing files	0	0.060801	1.0	0.23
Standardizing genes	0	1.0	0.769104	0.23



Raw data.

Shahriyari L. (2018) Briefing in Bioinformatics

Standardized data. In this figure, top panels (A–C) represent the results of independently standardizing each file, bottom panels (D–F) represent the results of independently standardizing each gene.



Briefings in Bioinformatics, Volume 20, Issue 3, May 2019, Pages 985–994, https://doi.org/10.1093/bib/bbx153

Shahriyari L. (2018) Briefing in Bioinformatics

OXFORD UNIVERSITY PRESS

Parameter estimation

- Number of equations is much smaller than the number of unknown parameters leading to infinite number of optimal values for parameters.
- Finding the structure of the space of optimal parameters' values for each patient.
- Looking at the intersection of spaces of optimal values across patients to reduce the dimension of the space of optimal regime.
- Collaborating with MPI Dr. Roy to perform parameter estimation and MPI Dr. Pal to perform sensitivity analysis.
- Generating test cases (generating artificial data using equations with some values for parameters), to evaluate the parameter estimation.
- Looking to collaborate with scientists in the field of Topology Optimization.

Optimal parameter estimation

- We will estimate optimal parameters by formulating and solving an optimal control problem.
- The control problem consists of minimizing a functional consisting of a least squares data-fitting term and regularization terms.
- The data-fitting term determines the optimal control parameter set which that leads to the best fit of our measured gene expression data.
- The regularization terms represent certain apriori assumptions on the parameter set.
- Such an optimization framework is robust and can accurately predict the optimal parameter set even with noisy or low number of measurements.
- This approach is new in the context of QSP models.

Sensitivity analysis and optimal treatment strategy

- It is important to understand the effects of model parameter values on the outcome measures.
- Sensitivity analysis is important because uncertainty in the obtained parameter values may result in variability in the model's prediction of resulting dynamics.
- To carry out the proposed sensitivity analysis, we will use cutting edge statistical tools like Partial Rank Correlation Coefficient Analysis and Latin Hypercube Sampling Scheme.
- For devising effective optimal treatment strategies, we will use a repetitive combination of non-linear optimization and global sensitivity analysis methods.

We are eagerly looking forward to your assistance

Please share with us your comments and insights on the project's GitHub page:

https://github.com/ShahriyariLab/Data-Driven-QSP-Software-for-Personalized-Colon-Cancer-Treatment

Eight open issues: QSP model, Tumor deconvolution, Data sets, Data preparation, Parameter estimation, Sensitivity analysis, Modeling treatment, and validation

Looking for Postdoc joining our team at UMass Amherst
lshahriyari@umass.edu

Thanks! Questions?