Quantitative Understanding of HT data in disease biology
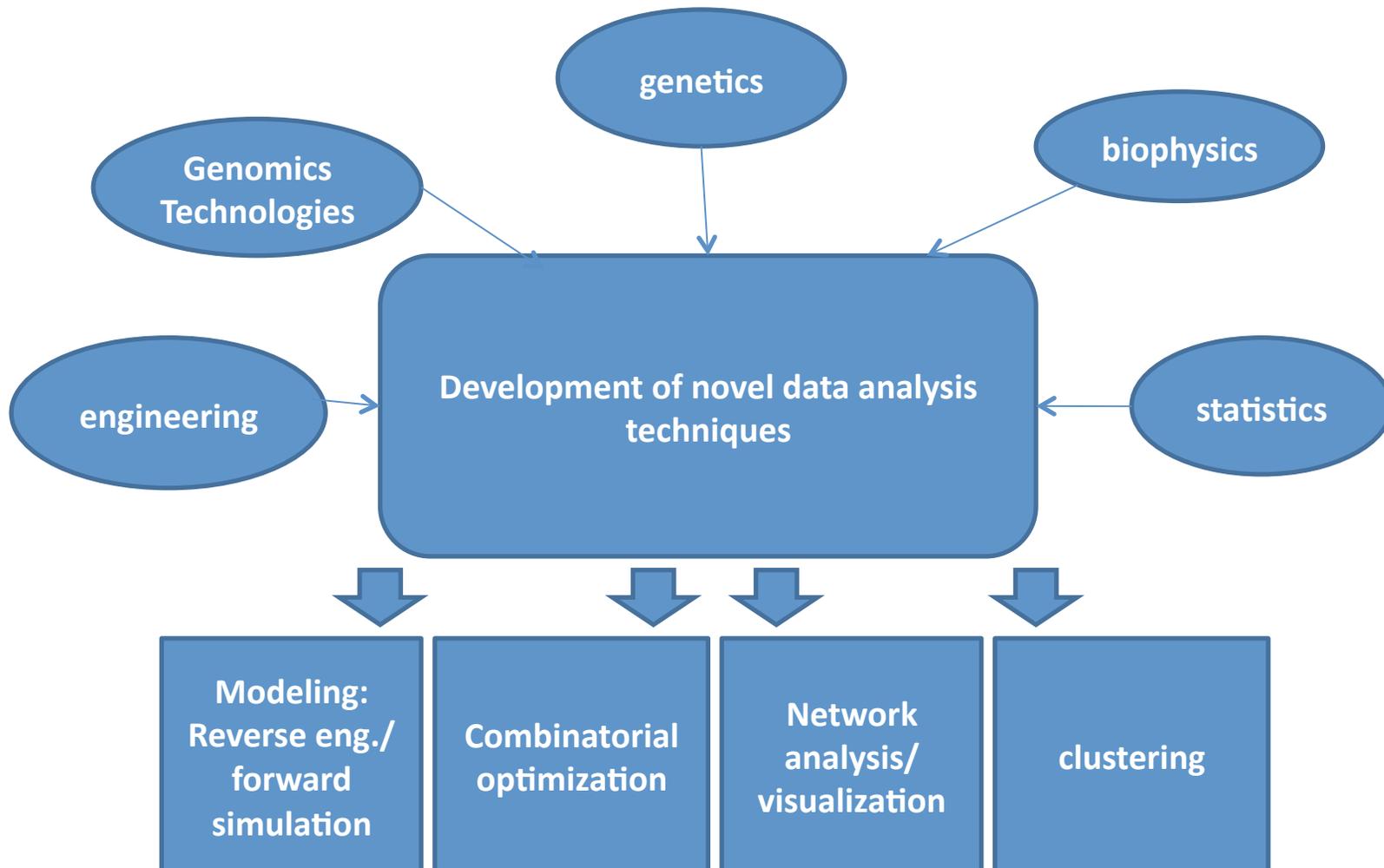
Zeynep H. Gumus, PhD

Institute for Computational Biomedicine and the Department of Physiology and Biophysics

HT analysis for personalized genomic medicine

- Generate, analyze and integrate various types of data to understand the mechanistic details of the underlying cellular processes in disease states.
  - **(**1) Development of novel HT data analysis techniques based on the concepts of biophysics, genetics, genomics, ngineering and statistics, including modeling, reverse engineering, network analysis and visualization and combinatorial optimization.
  - (2) Integrative analysis of various –omics and other HT databases using optimization and machine learning technologies to identify disease-specific features, such as biomarkers.
  - (3) Utilization of our or other methods to address mechanisms in cancer and neurodegenerative disease.

- In specific areas, such as in neurodegeneration, the goal of the cumulative effort is to find causal changes related to disease pathogenesis, and identify therapy targets that lead to the design of therapeutic methods and their implementation in personalized medicine. Within this framework, I will present how, fitting model parameters to data, sensitivity and robustness tests of ODEs fit into our research.

HT Data analysis techniques in our lab:

HT analysis for personalized genomic medicine

- Generate, analyze and integrate various types of data to understand the mechanistic details of the underlying cellular processes in disease states.
    - (1) Development of novel HT data analysis techniques based on the concepts of biophysics, genetics, genomics, ngineering and statistics, including modeling, reverse engineering, network analysis and visualization and combinatorial optimization.
    - **(2) Integrative analysis of various –omics and other HT databases using optimization and machine learning technologies to identify disease-specific features, such as biomarkers.**
    - (3) Utilization of our or other methods to address mechanisms in cancer and neurodegenerative disease.

- In specific areas, such as in neurodegeneration, the goal of the cumulative effort is to find causal changes related to disease pathogenesis, and identify therapy targets that lead to the design of therapeutic methods and their implementation in personalized medicine. Within this framework, I will present how, fitting model parameters to data, sensitivity and robustness tests of ODEs fit into our research.

Integrative data analysis

The Cancer Genome Atlas — *Understanding genomics to improve cancer care*

- A massively large scale multi-institute undertaking **to increase our understanding of the molecular basis of cancer** to aid diagnosis, treatment and prevention.
- Started in 2005; National Cancer Institute and National Human Genome Research Institute selected participating people and laboratories.
- Unique in (i) the number of patients (500/cancer, 500/normal) and (ii) number of high-throughput genomic techniques used to analyze patient samples:
  - **(~80% cases)**
  - **Gene expression profiling**
  - **miRNA expression profiling**
  - **DNA methylation profiling**
  - **SNP genotyping**
  - **Copy number variation profiling**
  - **Exon sequencing**
  - **Whole genome-sequencing (some tumors)**
  - **whole exon-sequencing (~80% of cases)**
  - **Phenotype information (age, gender, tumor stage)**
- 20-25 tumor types
- In FY 2010, new centers funded to characterize these tumors:
  - **Genome Characterization Centers (GCCs)**
  - **Genome Data Analysis Centers (GDACs) *dedicated funding to bioinformatics**

http://cancergenome.nih.gov/

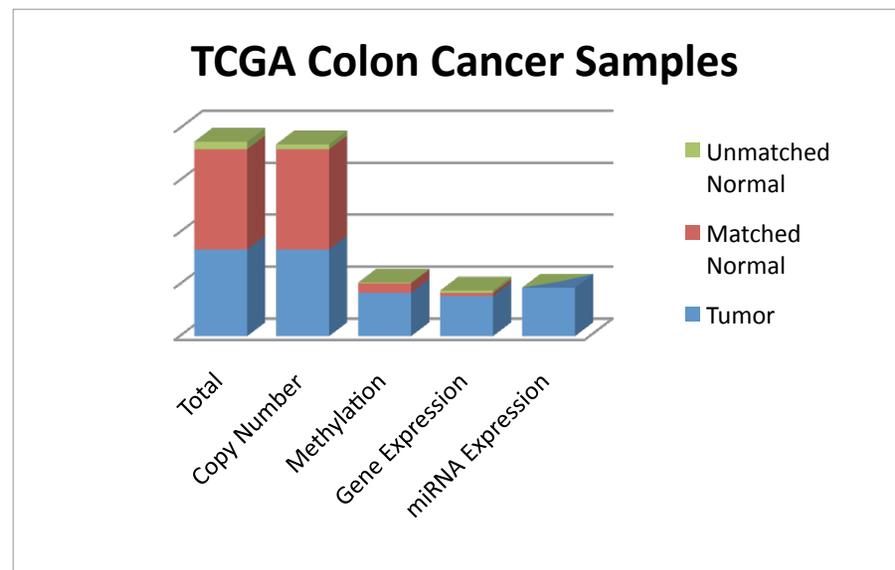# TCGA Objective: Identify all the changes in DNA of cancer cell. Why?

- **TCGA laying the foundation for Personalized Medicine**.

- Catalogs all the changes in the genomes and epigenomes of many samples
- Data from several different technology platforms, producing different type of genomic information
  - single base pair changes, extra copies or deletion of genes, increased or decreased gene expression, changes in chemical marks on DNA (methylation)
- **Specific genomic changes connected to specific outcome helps develop more effective, individualized therapies**
- Changes may reflect
  - **affected areas of the genome**
  - **functions/pathways controlled by those areas**
  - **What drugs can target them?**
- Or a change is linked to
  - **how fast the disease progresses or comes back after treatment**
  - **Disease subtype, patient phenotype**
  - **Specific signatures may respond differently to various treatments**

- Many steps needed to translate genomic data to patient care:
  - **which genomic changes are truly responsible for the disease**
  - **identifying or developing therapies to correct the impact of those changes.**

- **This type of Personalized Medicine is not yet common, but studies like TCGA will make it a reality in the future.**
- **We are utilizing the TCGA data to identify biomarkers for Colon Cancer prognosis.**

The Cancer Genome Atlas

Understanding genomics to improve cancer care

COAD	Total	Copy Number	Methylation	Gene Expression	miRNA Expression
Tumor	333	333	166	154	186
Matched Normal	386	386	37	12	0
Unmatched Normal	29	19	3	9	0

- 423 Patients with samples
- 333 Downloadable Tumor Samples

(Last Update 09/02/11)

TCGA Colon Cancer Samples

- Unmatched Normal
- Matched Normal
- Tumor

Total · Copy Number · Methylation · Gene Expression · miRNA Expression

http://cancergenome.nih.gov/

HT analysis for personalized genomic medicine

- Generate, analyze and integrate various types of data to understand the mechanistic details of the underlying cellular processes in disease states.

  - (1) Development of novel HT data analysis techniques based on the concepts of biophysics, genetics, genomics, ngineering and statistics, including modeling, reverse engineering, network analysis and visualization and combinatorial optimization.

  - (2) Integrative analysis of various –omics and other HT databases using optimization and machine learning technologies to identify disease-specific features, such as biomarkers.

  - **(3) Utilization of our or other methods to address mechanisms in cancer and neurodegenerative disease.**

- In specific areas, such as in neurodegeneration, the goal of the cumulative effort is to find causal changes related to disease pathogenesis, and identify therapy targets that lead to the design of therapeutic methods and their implementation in personalized medicine. Within this framework, I will present how, fitting model parameters to data, sensitivity and robustness tests of ODEs fit into our research.
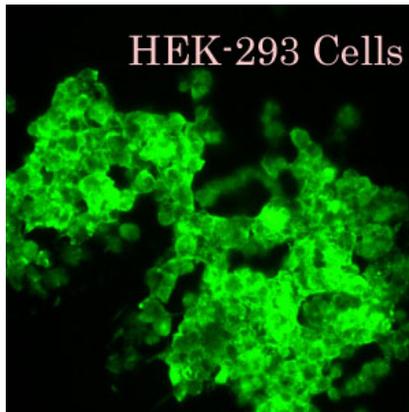
Experimental and computational studies at various scales

- Molecular
- Cellular
- Organismal
- Mechanisms of hallucinogenic drugs of abuse?
- Mathematical models of receptor-mediated signaling properties used to connect to experimentally determined signaling.
- Molecular complexes and interactions of these compounds with serotonin receptors in the G protein coupled receptors GPCRs family and follow the mechanisms through ensuing interaction processes with other components of signaling cascades such as membrane proteins (e.g. PIP2) and PDZ domains.

# Example: What are the signaling mechanisms of hallucinogenic drugs of abuse?

serotonin

HEK-293 Cells

www.altogen.com

- **Hallucinogens:** psychedelics, dissociatives, and deliriants.
- **Models necessary to develop a clear understanding of signal propagation into the cell.**
- **Example: experimental & computational study to understand the design principles of cell signaling mechanisms of hallucinogens binding to serotonin receptors that cause ERK1/2 activation. (Chang, Gumus, Weinsten 2009).**

- **Computational:** develop a mathematical model of receptor-mediated ERK1/2 activation in cells expressing 5-HT1A, 5-HT2A subtype serotonin receptors individually and together.

- **Experimental:** measure activation of ERK1/2 by action of selective agonists on these receptors.

- Investigation performed in HEK293 cells originally derived from Human Embryonic Kidney cells grown in tissue culture (source a healthy aborted fetus).

- HEK293 are ideal in investigation of protein-protein interaction: (i) extremely easy to culture and to transfect; (ii) used simply as a test tube with a membrane, where the cell behavior itself is not of interest..
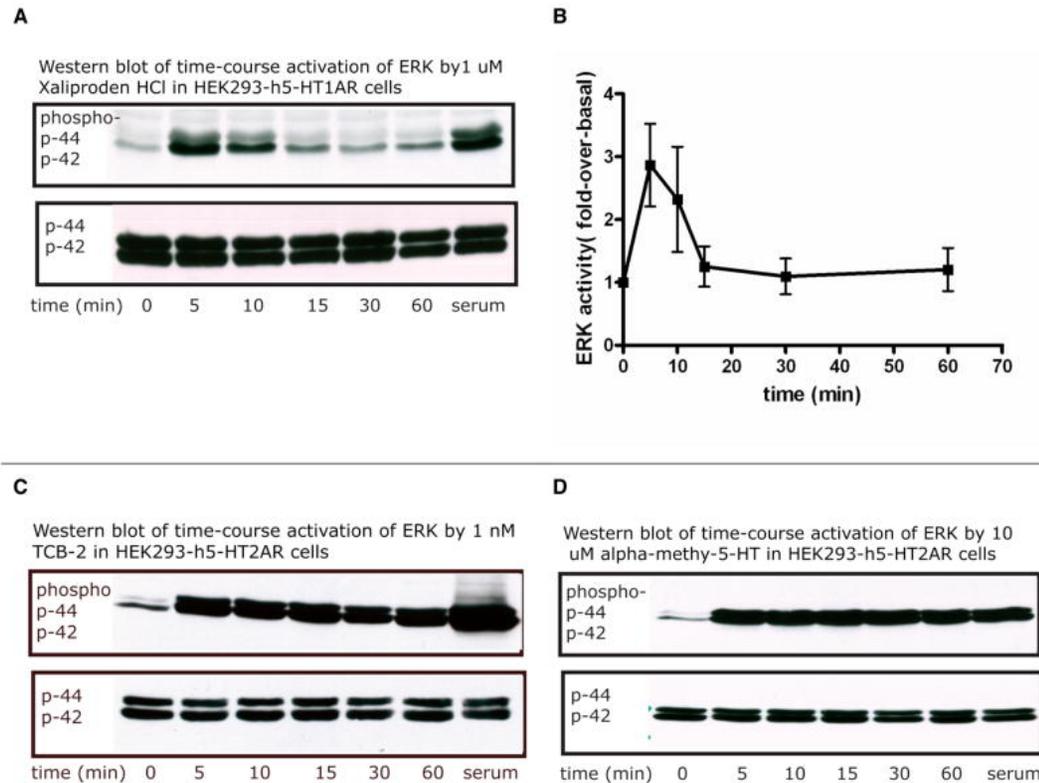
MAPK cascade activation by serotonin receptors: topological network (5-HT1A and 5-HT2A)
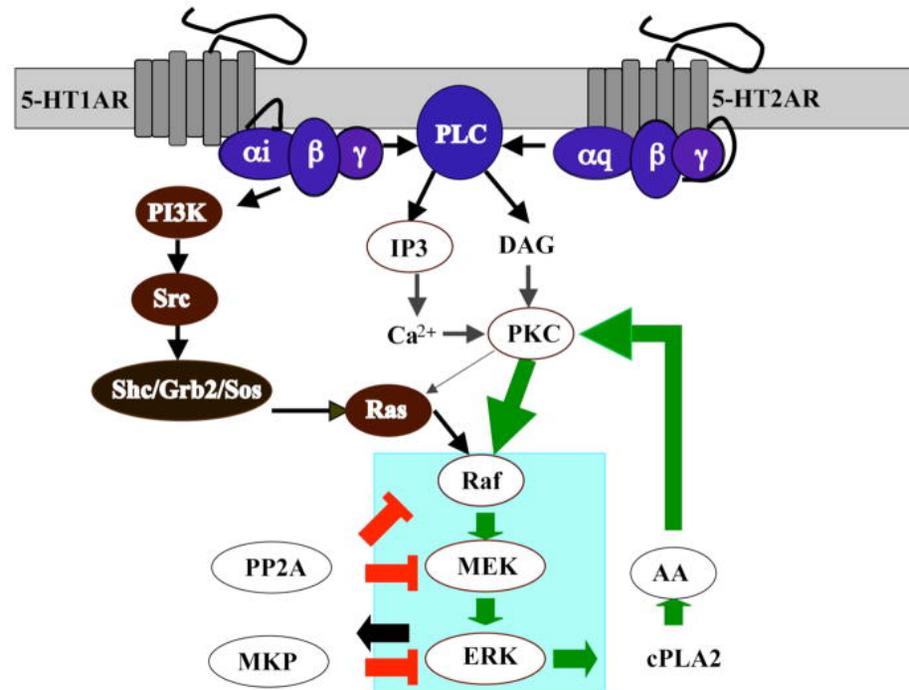
- MAPK cascade ubiquitously expressed in diverse biological processes
- MAPK signal transduction pathways mediate
  - short-term effects (modulation of potassium channel and glutamate receptor function)
  - Long-term effects (cell differentiation, long-term potentiation, learning and memory)
  - Signaling through MAPK pathway positively regulates immediate early genes
  - Tightly regulated in a variety of cells by multiple feedback loops
  - basic structure of all MAPK cascades is the same, but differences in feedback control enable them to generate a plethora of biological responses, including oscillations, gradual and ultrasensitive responses
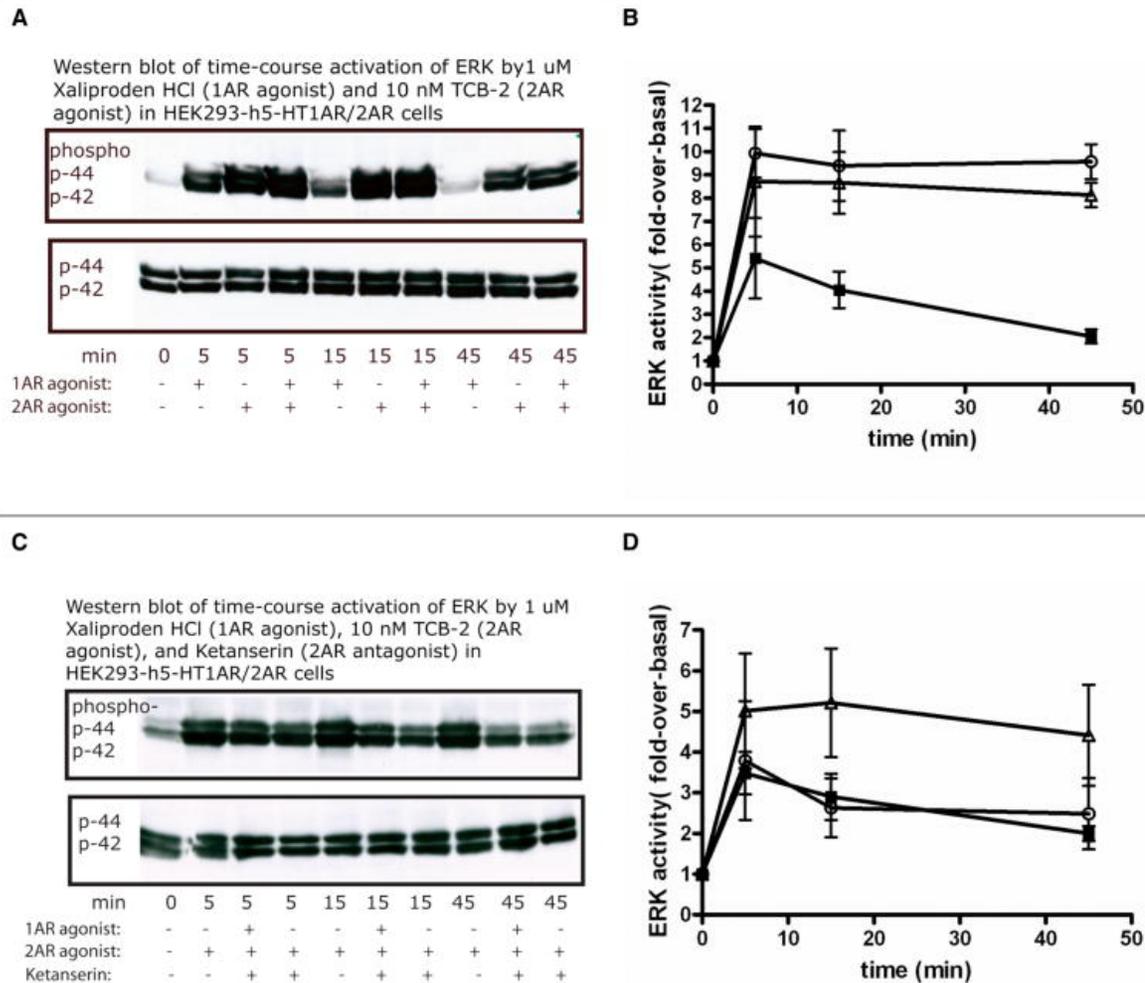
Why do modeling?

- 5-HT2A receptors… direct targets of hallucinogens… but balance of signaling activities that produce hallucinogenic effect remains unknown.

- **How are activated receptors integrated into signaling pathways?**

- How can specific conformations of the activated receptor establish the distinct patterns of signal transduction observed when they bind different ligands?

- E.g. hallucinogens produce entirely different transcriptome fingerprints compared to non-hallucinogenic congeners.

- **Which of the reactions in these complex networks are important?**

- **Where are the cross-talk points regulated by upstream, downstream components?**

- Essential to acquire qualitative information on which interactions take place and quantitative data on their strength and modeling to integrate such information.

A

Western blot of time-course activation of ERK by1 uM Xaliproden HCl in HEK293-h5-HT1AR cells

phospho-
p-44
p-42

p-44
p-42

time (min)   0   5   10   15   30   60   serum

B

C

Western blot of time-course activation of ERK by 1 nM TCB-2 in HEK293-h5-HT2AR cells

phospho
p-44
p-42

p-44
p-42

time (min)   0   5   10   15   30   60   serum

D

Western blot of time-course activation of ERK by 10 uM alpha-methy-5-HT in HEK293-h5-HT2AR cells

phospho-
p-44
p-42

p-44
p-42

time (min)   0   5   10   15   30   60   serum

- Representative Western blot (A, C, D) and quantitative analysis (B) of time-course activation of MAPK in HEK293 cell lines, stably expressing only human 5-HT1A or 5-HT2A receptors. Note that B is the quantitative result of three experiments in HEK293-h5-HT1AR cells with 1µM of Xaliproden HCl (shown by solid squares). Serum (Bovine Calf Serum) was used as the positive control. C and D show experiments in which the HEK293-h5-HT2A receptors were treated with 1nM TCB-2 and 10µM α-methyl-5-HT, respectively. Both C and D indicate that 5-HT2AR activation resulted in a sustained ERK activation.

ERK-activation

- Xaliproden HCl: 5-HT1AR agonist.
- TCB-2: high-affinity 5-HT2AR agonist.
- Ketanserin: 5-HT2AR antagonist.

A

Western blot of time-course activation of ERK by1 uM Xaliproden HCl (1AR agonist) and 10 nM TCB-2 (2AR agonist) in HEK293-h5-HT1AR/2AR cells

phospho
p-44
p-42

p-44
p-42

min	0	5	5	5	15	15	15	45	45	45
1AR agonist:	-	+	-	+	+	-	+	+	-	+
2AR agonist:	-	-	+	+	-	+	+	-	+	+

B

ERK activity(fold-over-basal) vs time (min)

C

Western blot of time-course activation of ERK by 1 uM Xaliproden HCl (1AR agonist), 10 nM TCB-2 (2AR agonist), and Ketanserin (2AR antagonist) in HEK293-h5-HT1AR/2AR cells

phospho-
p-44
p-42

p-44
p-42

min	0	5	5	5	15	15	15	45	45	45
1AR agonist:	-	-	+	-	-	+	-	-	+	-
2AR agonist:	-	+	+	+	+	+	+	+	+	+
Ketanserin:	-	-	+	+	-	+	+	-	+	+

D

ERK activity(fold-over-basal) vs time (min)

- Time-course activation of ERK in HEK293 cell lines, stably expressing both human 5-HT1A and 5-HT2A receptors. Representative Western blots are shown in A, C and quantitative analyses of three experiments are displayed in B, D. The agonists Xaliproden HCl (1 µM) and TCB-2 (10 nM) were used to activate 5-HT1A and 5-HT2A receptors, respectively, in the presence (C, D) or absence (A, B) of 50nM Ketanserin. In panel B, solid squares denote 5-HT1A-specific response; open triangles denote the 5-HT2A response; and open circles denote the combined response of the two receptors. In panel D, solid squares denote the combined response of 5-HT1A and 5-HT2A in the presence of Ketanserin; open triangles denote the 5-HT2A-mediated ERK activation; open circles denote the same response in the presence of Ketanserin.

MAPK cascade activation by serotonin receptors: mathematical model (5-HT1A and 5-HT2A)
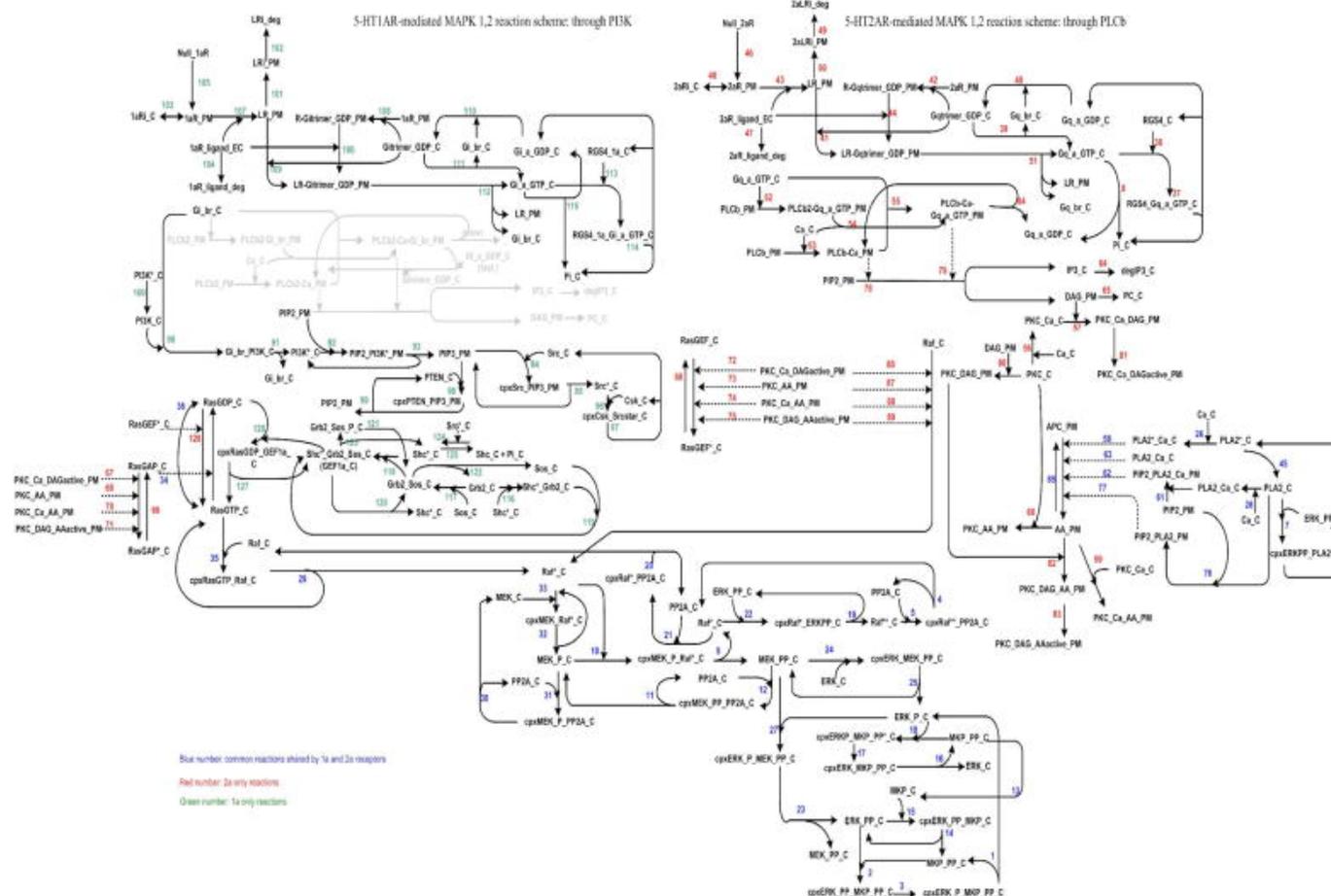
Integrative representation of currently known kinetics.
light grey area is cell-type specific/not demonstrated in HEK293.
Network representations are incomplete; parameters have significant uncertainties
* Model simulations can aid ongoing efforts to construct an increasingly comprehensive mechanistic understanding by validating/eliminating specific assumptions, answering particular questions and guide experiments by producing testable hypotheses.

The mathematical model & simulation

- 112 species
- 228 parameters
- 128 real reactions
- Simulate at
  - (i) steady state (no concentration changes prior to stimulation)
  - Time course (variables are determined as at time series upon ligand treatment, using the steady-state concentrations as starting values)
  - Simulate each receptor alone, and together.
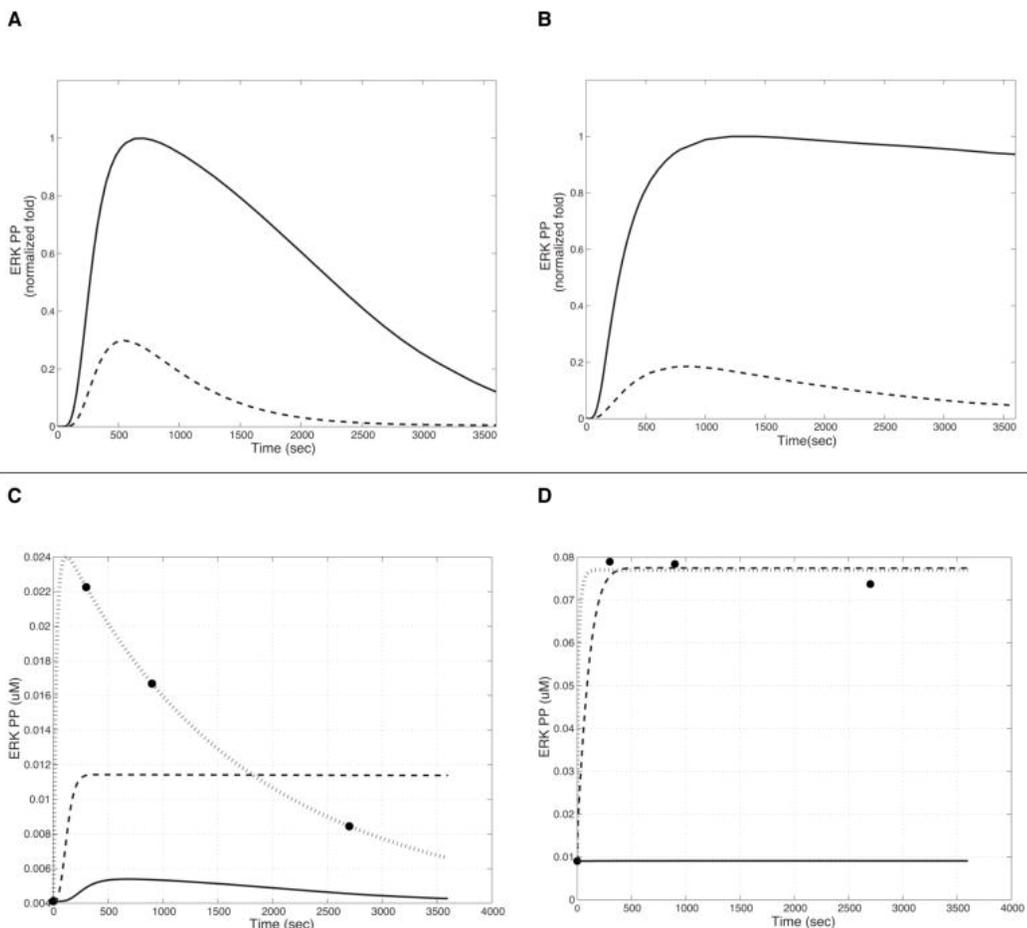
**Simulation of time-course MAPK activation in cells stably expressing human 5-HT1A and 5-HT2A receptors**.
A: Phosphorylated ERK (ERK_PP) 5-HT1AR stimulation by ligands w/ different efficacy/ receptor activity values.
B: ERK_PP after stimulation of 5-HT2AR by ligands w/ 2 different efficacy/receptor activity values.
C & D: results of parameter optimization. Circles-average experimental values from three experiments in cells expressing both receptors, stimulated with either 1A agonist (in C) or 2A agonist (D). In both, solid- initial results, interrupted - result from optimization of the most sensitive parameter (k51 in C; k82 in D). Dotted- results of simulation with optimized parameters (all parameters with $|S| > 0.3$ from the sensitivity analysis). Final R^2 values for this simulation are 0.983 in 6C, and 0.995 in 6D.

Signal transmission mechanisms

- (i) protein protein interactions and enzymatic reactions such as protein phosphorylation and dephosphorylation
- (ii) protein degradation or production of intracellular messengers.
- Basic chemical reaction schemes:

$$A + B \xrightleftharpoons{k_f, k_b} AB$$

$$A + B \xrightleftharpoons{k_f, k_b} C + D$$
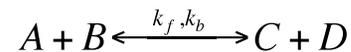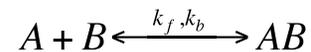
- The reactions are completely specified by rate constants and initial concentration of each reactant (A, B, C, D).
- How do you estimate the parameters?
  - published experimental studies
  - Optimization
  - The mathematical/kinetic pathway model should yield results very similar to experimental data

Sensitivity analysis

- To capture essential model characteristics
- To determine critical parameters that control peak phosphorylation of ERK1/2:
  - Initial conditions of species states
  - Kinetic rate constants
  - Plot time-dependent sensitivities of phosphorylated ERK1/2 to changes in parameters and identify sensitivity values at the time point of phosphorylated ERK1/2 peaks
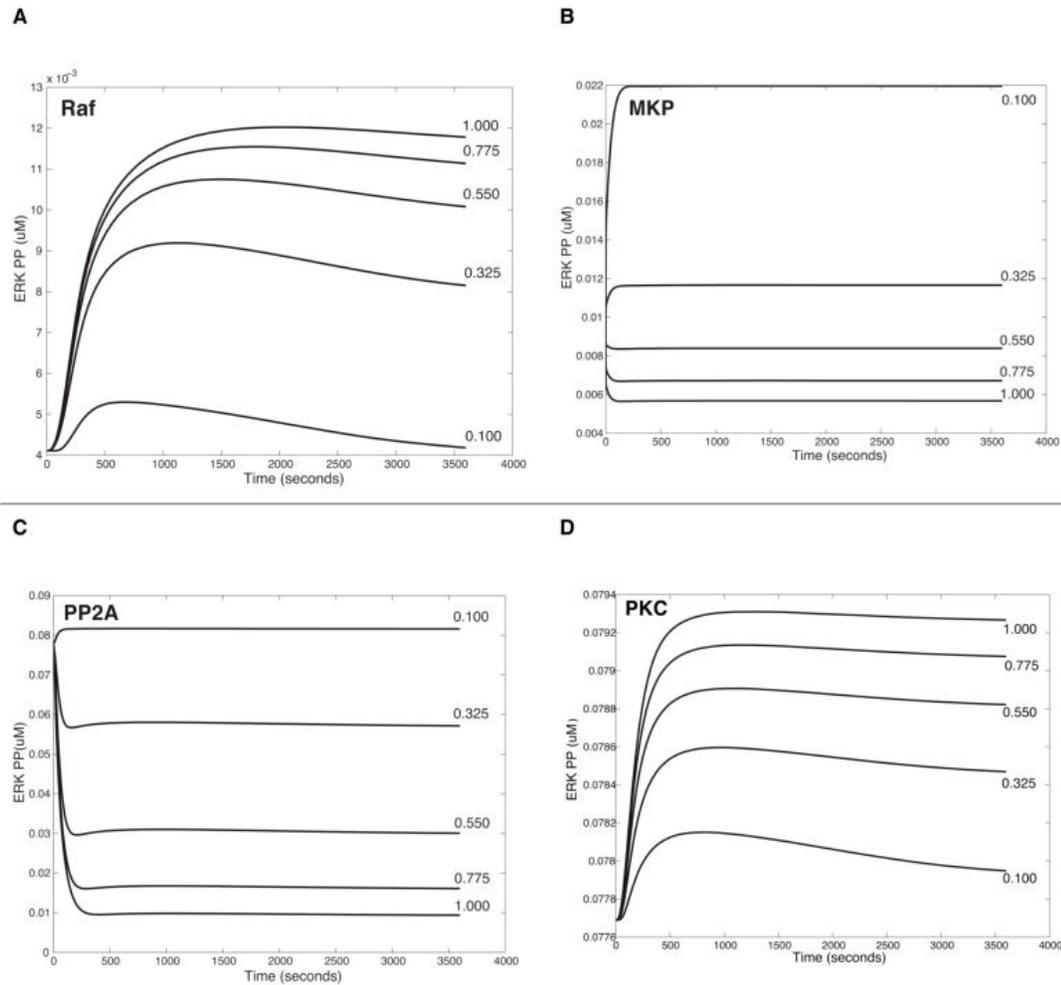
- *In silico* experiment of parameter perturbation for state variables ERK activation. A. Increase of Raf in the 5-HT1AR-triggered pathway, B. Variation of MKP in the 5-HT2AR-triggered pathway; C &D. Changes of PP2A (in C) and of PKC (in D) in the combined pathways triggered by the two receptors (5- HT1A/2AR). Note that concentrations of ligands in models are as used in the experiment: 1 µM Xaliproden HCl for the 5-HT1A receptor; 10 nM of TCB-2 for the 5-HT2A receptor.

Parameter Estimation

- **Mathematical model: a set of coupled ODEs; parametric reaction rate constants; initial concentrations.**

- **Can modifying critical parameters lead to a better model fit of experimental ERK1/2 data?**

  - Min. sum of squared errors between experimental and simulated phosphorylated ERK1/2 concentration at different time points by manipulating these parameters.

  - Assume experimental error normally distributed with zero mean and covariance matrix same in each time point and diagonal.

  - Identify a set of parameter combinations that lead to observed phosphorylated ERK1/2 behavior.

  - **Best parameter combination is not unique: different sets of combinations can lead to the same observed output, an inherent property of multi-parameter nonlinear systems biology models.**

$$\min_{\mathbf{k}, \mathbf{y}_\mu} \sum_{\mu=1}^{4} (ERK_{PP,\mu} - \overline{ERK}_{PP,\mu})^2$$

$$\text{s.t.} \frac{dy_i}{dt} = g_i(\mathbf{y}, \mathbf{k}, \mathbf{t}), \forall i \in I$$

$$y_i(t_0) = \mathbf{y}_{0i}, \forall i \in I$$

$$\mathbf{t} \in [t_0, t_f], \mathbf{y} \in [\mathbf{y}^L, \mathbf{y}^U]; \mathbf{k} \in [\mathbf{k}^L, \mathbf{k}^U]$$

$$\mathbf{ERK_{PP}} \in \mathbf{y}.$$

i -index of components with explicit time derivatives in the model of size I,
y -vector of time-dependent species concentrations (state variables);
y0- initial concentration of at t0;
K- vector of rate constants to be estimated;
ERK_PPM- experimental and simulated values of phospho-ERK1/2 at m[th] data point,
Function **g** involves changes of species upon time and rate constants.
Four experimental data points of phospho-ERK1/2: 0, 5, 15, 45 minutes, each based on the average of three experimental results. Lower and upper bounds of **y** and **k** are indicated by *L and U,* respectively.

Convex objective function , but the reaction rate terms in constraints **g** are **nonlinear and nonconvex**; **problem is nonconvex**. Mminimization is over both the rate constants (**k**) and the state variables (**y**), this is true for constraints involving even the simplest linear reaction rate term , which is bilinear **(nonconvex).**
Solution of this nonconvex problem with local optimizers leads to a local optimum.

Simulations and Optimization Algorithm

- Simulations carried out using SimBiology Toolbox within MATLAB

- Differential equations were integrated using ode15s function, which as a variable order solver, based on the numerical differentiation formulas and designed for stiff systems

- Local optimization performed using fmincon and lsqnonlin functions in SimBiology toolbox

Mechanisms of signal transfer

- Regulation by protein-protein interactions
- Protein phosphorylation
- Regulation of enzymatic activity
- Production of second messengers
- Cell surface signal transduction systems

Networking results in persistent activation of protein kinases after transient stimulus– implicated in diverse processes such as neoplastic transformation and learning and memory. Mutations or altered gene expression can lead to persistent activation.

Do connections between pre-existing signaling pathways result in persistently activated protein kinases capable of eliciting end-point biological effects?

Conclusions

- Detailed topological representation of 5-HT1A and 5-HT2A receptors-mediated ERK activation
- Based on known reactions and assumptions derived from canonical pathways
- Translate into mathematical equations that describe the network topology
- Computationally simulate the equations
- Lead to predictions of species concentration profiles that vary with respect to time upon ligand stimulation
- Cells that express both receptors produce dynamics distinct from receptors that are expressed alone.
- When both are expressed, 5-HT2A dominates the ERK signal.
- Treatment with 5-HT2A antagonist (Ketanserin) produces a switch in ERK activation pattern from sustained to transient.
- Individual pathway models produce results in qualitatively good agreement with the experimental data, and parameter optimization establishes quantitative agreeement.

Conclusions

- Important parameters and intrinsic behaviors of the system are further revealed by sensitivity analysis and parameter optimization.

- Models suggest that constitutive activity combined with specific drug efficacy may determine distinct dynamics of 5-HT receptors mediated ERK1/2 pathway, affecting the receptor activation phenotypes.

- *In silico* experiments provide insights to the underlying mechanisms of ERK pathways via 5-HT receptors which can be further valided by inhibitors or activators, siRNA or trasfections to influence the activity and expression of target genes.