A Framework for Classification of Heterogeneous Calcium Dynamics

Rishikesh Kumar Gupta

(CH13M1011)

A Thesis Submitted In Partial Fulfillment of the Requirements
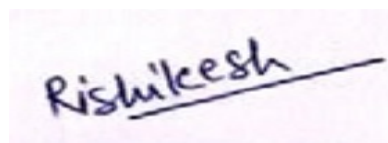
For The Degree of

Master of Technology

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Department of Chemical Engineering

Indian Institute of Technology Hyderabad

June, 2015

Declaration

I declare that this written submission represents my ideas in my own words, and where others' ideas or words have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that any violation of the above will be a cause for disciplinary action by the Institute and can also evoke penal action from the sources that have thus not been properly cited, or from whom proper permission has not been taken when needed.
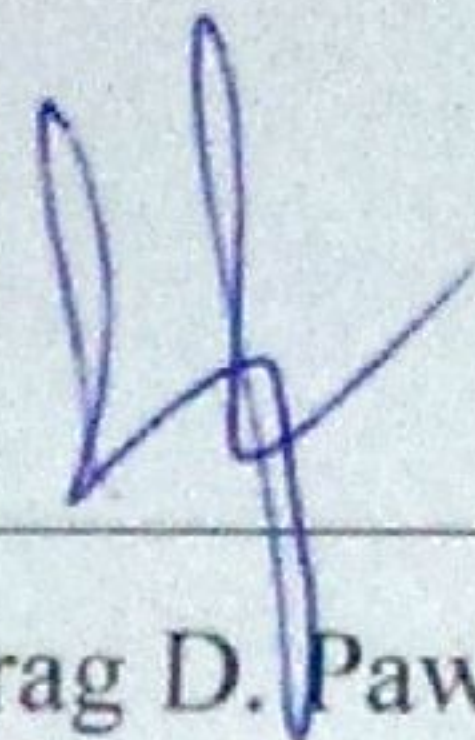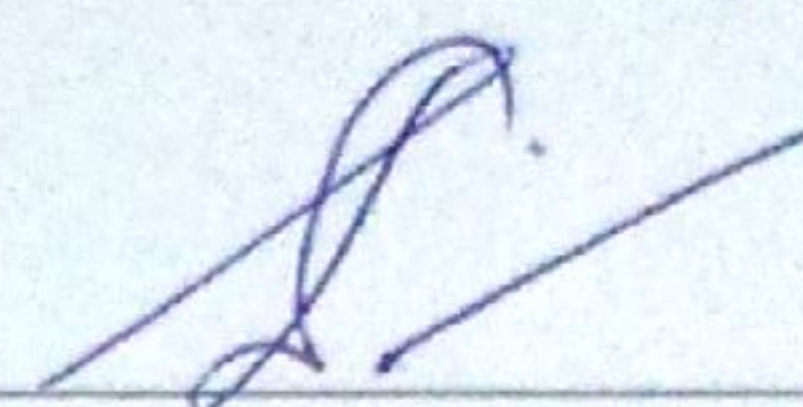
(Signature)
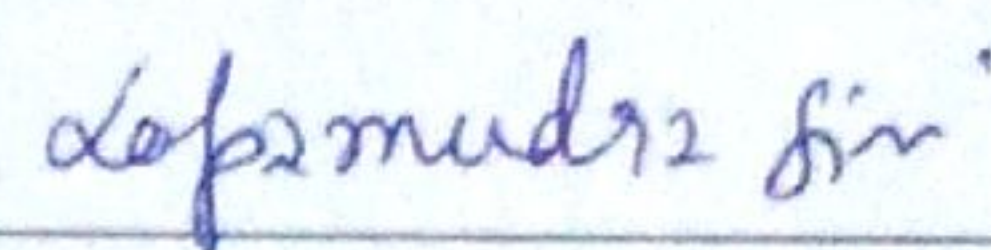
Rishikesh Kumar Gupta

(CH13M1011)

Approval Sheet

This thesis entitled **A Framework for Classification of Heterogeneous Calcium dynamics** by **Rishikesh Kumar Gupta** is approved for the degree of Master of Technology from Indian Institute of Technology-Hyderabad.
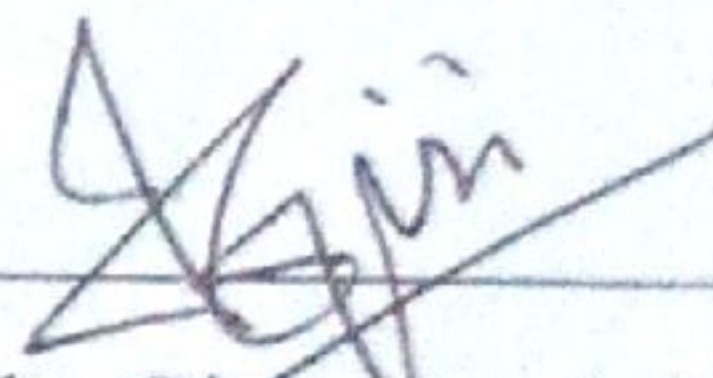
(Dr. Parag D. Pawar) Examiner

Dept. of Chemical Engineering

IIT-Hyderabad

(Dr. Saptarshi Majumdar) Examiner

Dept. of Chemical Engineering

IIT-Hyderabad

(Dr. Lopamudra Giri) Adviser

Dept. of Chemical Engineering

IIT-Hyderabad

(Dr. Jyotsnendu Giri) Co-Adviser

Dept. of Biomedical Engineering

IIT-Hyderabad

iii

Acknowledgements

I thank almighty whose blessing has enabled me to accomplish my project work successfully.

It is my pride privilege to express my sincere thanks and deep sense of gratitude towards **Dr. Lopamudra Giri**, for her valuable advice, fantastic supervision and constant patience through which this work was able to take a shape in which it has been presented. It was her valuable discussions and endless endeavours through which I gained a lot. Her constant encouragement and confidence-imbibing attitude has always been a moral support for me. Again I am very thankful to him for her encouragement, her time and engagement, and for sharing her wide knowledge with me.

A special word of thanks to my thesis evaluation committee Dr. Parag D. Pawar, Dr. S. Majumdar, Dr. J. Giri for their valuable suggestions.

I feel lacunae of words to express my most heartfelt gratitude to my friends **Mr Sarpras Swain** (Ph. D. Scholar) who have always been a source of inspiration for me and stood by my side at the toughest times.

On a more personal note, the whole credit of my achievements goes to my parents, who were always there for me in my difficulties. It was their unshakable faith in me that has always helped me to proceed further.

Rishikesh Kumar Gupta

Dedicated to My parents.

Abstract

G-protein coupled receptors are the target receptors for designing almost 45% of the drug in current drug market. For drug screening using cell-based assays, fluorescent imaging of GPCR mediated calcium dynamics in single cells can be used to obtain the dose-response profile. However, construction of a dose-response function based on single cell responses is rather challenging as the cells in a population respond heterogeneously to the drug. Here we developed a live cell imaging-based approach to quantify the heterogeneity of the HeLa cell population in response to GPCR mediated drugs. First, we found that activation of CXCR4 by SDF-1α induces calcium responses in Hela cells. We measured the temporal dynamics of cytosolic calcium through time-lapse imaging using confocal microscopy for various drug doses. We then reduced the high dimension data using principal component analysis and performed classification of temporal calcium dynamics using K-means algorithm. Time course of calcium concentration were found to be heterogeneous and such cell-to-cell variability was modeled as a mixture of three subpopulations. Using this sub-population model, we characterized the dose-response characteristics corresponding to CXCR4-mediated calcium responses at different dose ranges. Our technique was also validated using other GPCR receptors and drugs such as endothelin and norepinephrine mediated calcium responses. We also investigated the effect of cell crowding on calcium oscillations using the proposed tool and found that cell-to-cell contact plays a major role in calcium oscillation. Our approach provides a data analytic framework for automated quantification of cell-to-cell variability in time-course of cytosolic calcium responses. Most importantly, it offers a mode of classification of the cells with respect to its dynamic properties. More generally, the strategy presented here can be used to quantify the redistribution of subpopulations for time-course of intracellular concentrations from a large pool of cells obtained through microscopy.

Key words: GPCR, CXCR-4/SDF-1α, Calcium, live cell imaging, cell-to-cell variability

List of the tables

List of the Figures

List of the abbreviations

Abbreviations	Full Form
Ca^{2+}	Calcium ions
CXCR 4	Chemokine receptor 4
RNA	Ribonucleic acid
CXCL 12	Chemokine ligand 12
SDF-1α	Stromal derived factoer-1α
PCA	Principle Components analysis
GPCR	G-protein coupled receptors
ER	Endoplasmic reticulum
GnRH	Gonadotropin-releasing hormone
HUVECs	Human Umbilical Vein Endothelial Cells
RTK	Receptor tyrosine kinase
PLC	Phospholipase C
FRET	Fluorescence resonance energy transfer
HEK293	Human embryonic kidney cells
CD184	Cluster of differentiation 184
CHOs	Chinese hamster ovary (CHO) cells
PBMCs	Peripheral blood mononuclear cells
IEC-6	Intestinal epithelial cell lines
Caco2	Human colon tumor cell line
AM	Amplitude modulation
FM	Frequency modulation
AFM	Amplitude and frequency modulation
ISI	Inter spike interval
EST	Exercise stress test
MEM	Minimum Essential Medium
HBSS	Hanks' Balanced Salt Solution
ED-CCD	Electron-multiplying charge-coupled devices
SVM	Support vector machine
GMM	Gaussian mixture model

Contents

Chapter 1: Introduction

1.1-Calcium Signaling:

Calcium Signaling (also called as Ca^{2+} encoding/calcium information processing) is an intracellular signaling pathway used by a number of cells to transfer, process and convert external information sensed by the cell. In sense of physiological fate of cell, the external information is often
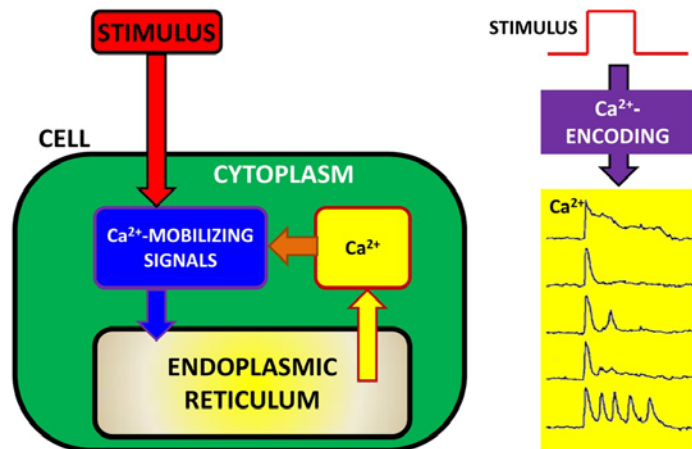
Figure 1.1: Calcium encoding

converted into intracellular calcium dynamics. The theory of calcium signaling describes how Ca^{2+} ions act as intracellular/secondary messengers, conveying information within cells to control their action. Given the role of Ca^{2+} ions in cell physiologies like cell migration, differentiation, contraction as well as metabolism, Ca^{2+} signaling has also been proposed as a forthcoming tool to depict cell functioning in diseased condition.

1.2-Why the heterogeneity in cell responses is important?
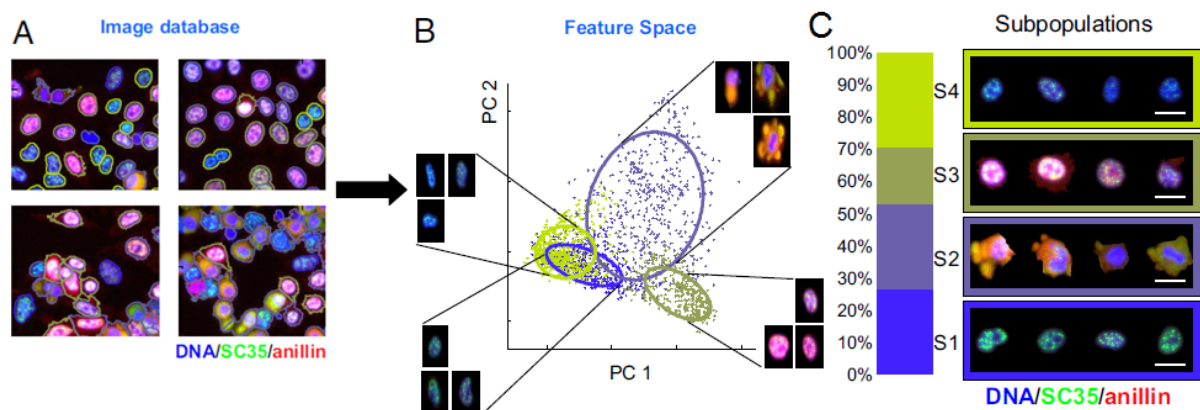
Figure 1.2: Heterogeneity in cell population and modeling as mixture of subpopulations

Observed heterogeneity in multicellular organisms can be categorized into four groups- (1) size and shape, (2) genome, (3) RNA expression levels, and (4) individual counts of metabolites. Behavior of multicellular organisms are determined by cell-type-specific combinations. A particular cell type may have identical genetic blueprint though they may have distinct tolerance and may respond in a diverse way to same environmental challenges.

Classically, the biological models (e.g., signaling networks and processing) have been conjectured based on experimental measurements of population-average behavior (e.g., mean, median or mode).
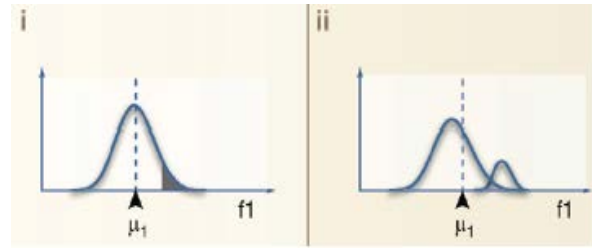
Figure 1.3: Mixture of cell population

However, given that the phenotypic cell-to-cell variability exists even between genetically homogeneous individuals and under identical stimulus, analysis and conclusions on cell responses is rather challenging.

The upcoming cell-response analysis focuses towards single-cell analysis and therefore inherent nature of cell-to-cell variability in a cell population has been identified. Single-cell gene expression analysis reveals an overlapping molecular signature (Figure 1.3) and its effect on cell-fate decisions by single cells. Such variability may carry valuable information that can facilitate the understanding of regulatory networks or the various cell states. The major challenge lies in determining the connection between such variability and the underlying biological significance. Figure 1.4 shows that epileptic condition shows high diversity pattern in neuron activity.
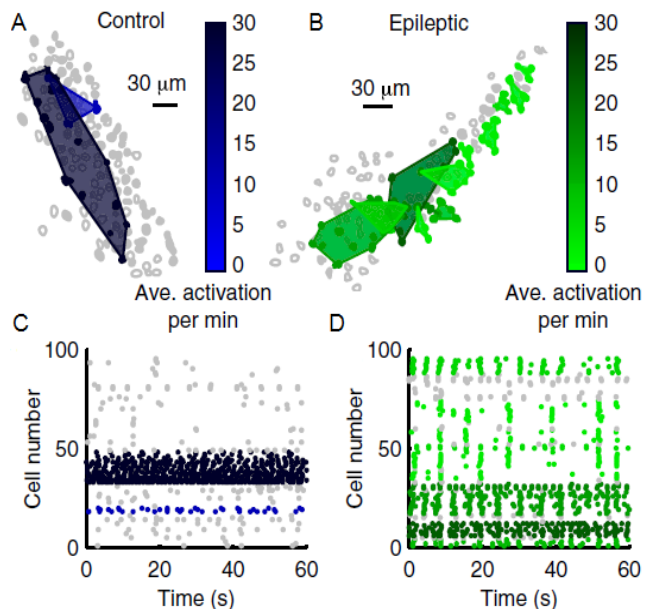
Figure 1.4: Cells classified on average frequency of calcium activation in neurons

To address this issue, development of new computational tools is needed, in synchronization with proper experimental platforms. This will lead to precise analysis of cell-cell variability and correlation with physiological responses.

1.3-Major reasons for origins of cell-to-cell variability

In a particular type of cell line, the number of receptor present on cell surface can be different that may lead to variability. Also, the kinetic parameters of cell physiology may be different for individual cells. Additionally, cell crowding effect/cell-cell contact could be different for each cells.

Chapter 2: Review Literature

2.1-Calcium:

Calcium is a chemical element with symbol

Ca and atomic number 20. Calcium is a soft gray alkaline earth metal, fifth-most-abundant element by mass in the Earth's crust. Ca is essential for living organisms, particularly in cell physiology, where transportation of the Ca ion in and out of the cytoplasm functions as a signal for many cellular processes.

The resting concentration of Ca^{2+} in the cytoplasm is normally maintained in the range of 10–100 nM. To maintain this low concentration, Ca^{2+} is actively pumped from the cytosol to the extracellular space and into the ER, and sometimes in the mitochondria. Certain proteins of the cytoplasm and other organelles act as buffers by binding Ca^{2+}. Signaling occurs when the cell is stimulated to release Ca^{2+} from intracellular stores, and/or when calcium enters the cell through plasma membrane ion channels (Clapham 2007).

2.2-Calcium signaling:

Calcium is a universal secondary messenger and plays a central role in controlling cell differentiation, proliferation and migration (Berridge, Bootman et al. 1998) (See Table 2.1, 2.2 and 2.3). Calcium encoding is a well-known mechanism to transfer the external information converted into intracellular calcium dynamics that eventually regulates the specific fate of the physiological process (Berridge, Lipp et al. 2000, Clapham 2007, Guo, Wang et al. 2015). Specifically many of the diseased condition is lined to the dysregulation of the calcium oscillation (Iino 2010, Vachel, Norez et al. 2015).

Table 2.1: Calcium signaling in cell cycle/cell division

Cell type/cell line	Techniques used	Main conclusion	References
Endosperm cells of *Haemanthus*	Fura-2 based fluorescent imaging	Ca^{2+} signals play an important role in controlling different phases of the cell cycle and interference in Ca^{2+} signaling may disrupt progression of the cell cycle	(Poenie, Alderton et al. 1986)
Yeast cells as model (somatic Cell)	Mathematical modeling	Calcium signaling triggers a cascade of events/reactions leading to the completion of mitosis.	(Baran 1996)

Table 2.2: Calcium signaling in cell migration and differentiation

Cell type/cell line	Techniques used	Main conclusion	References
GnRH neuron (neuron cell in brain)	Differential interference contrast microscopy, Migration assay, Immunostaining	The regulation of movement through Ca^{2+} signals that act on cytoskeletal dynamics. (migration)	(Hutchins, Klenke et al. 2013)
HeLa cells	Mathematical modeling	Circulating Gal-3 activates the Calcium sensitive PKC-dependent pathway by activating ERK1/2 promoting cell migration.	(Gao, Balan et al. 2014)
HUVECs	Confocal microscopy	Ca^{2+} acts as a secondary massager in RTK and PLC signaling.	(Tsai, Seki et al. 2014)

Table 2.3: Calcium imaging in drug screening and disease diagnosis

Target Disease	Techniques used	Cell line	Main Aim	References
Epilepsy (disease diagnosis)	Electrophys-iological Recordings	Stomato-sensory cortex	Morphological characterization of neurons and determine the single cell property.	(Badea, Goldberg et al. 2001)
Alzheimer's disease (drug screening)	FRET based confocal imaging (calcium monitoring)	HEK293	Establishment of high-throughput drug screening assay through calcium imaging.	(Honarnejad, Daschner et al. 2013)

2.3-Significance of GPCR mediate calcium response:

Since GPCR is the target receptor for almost 45% of the drug in pharmaceutical industry, quantification of GPCR mediated (Aggarwal, Sethi et al. 2007) calcium response is crucial in case of cell-signaling studies as well as drug screening (Honarnejad, Daschner et al.

2013). Screening of GPCR targeting drugs based on calcium imaging has also been reported (Table 2.3). Although such high resolution calcium imaging using confocal microscopy is becoming important for screening of drugs, the data analysis remains as a challenge. Pharmacokinetic or pharmacogenomics studies required for drug validations are based on the dose response characteristics (Michelini, Cevenini et al. 2010). Construction of such a dose-response curve based on single cell calcium dynamics is rather challenging as the cells show high cell-to-cell variability in their response to the drug.

2.4-Chemokine receptors 4(CXCR4): Structural integrity with GPCR

CXCR4 (also known as fusin, CD184) is a seven-pass transmembrane Class1 GPCR (G_i) (rhodopsin like GPCR family). It is a 352-355 amino acid protein span through plasma membrane with three extracellular (N-terminal) domain and three intracellular (C-terminal) domain (See Figure 2.1) (Busillo and Benovic 2007, Wegner, Ehrenberg et al. 1998). Here we choose to study calcium response mediated by the chemokine receptor type-4 (CXCR-4), that belongs to the GPCR family (Choi, Duggineni et al. 2012) and binds with its ligand stromal-cell derived factor-1α (SDF-1α) (Zlotnik and Yoshie 2012).
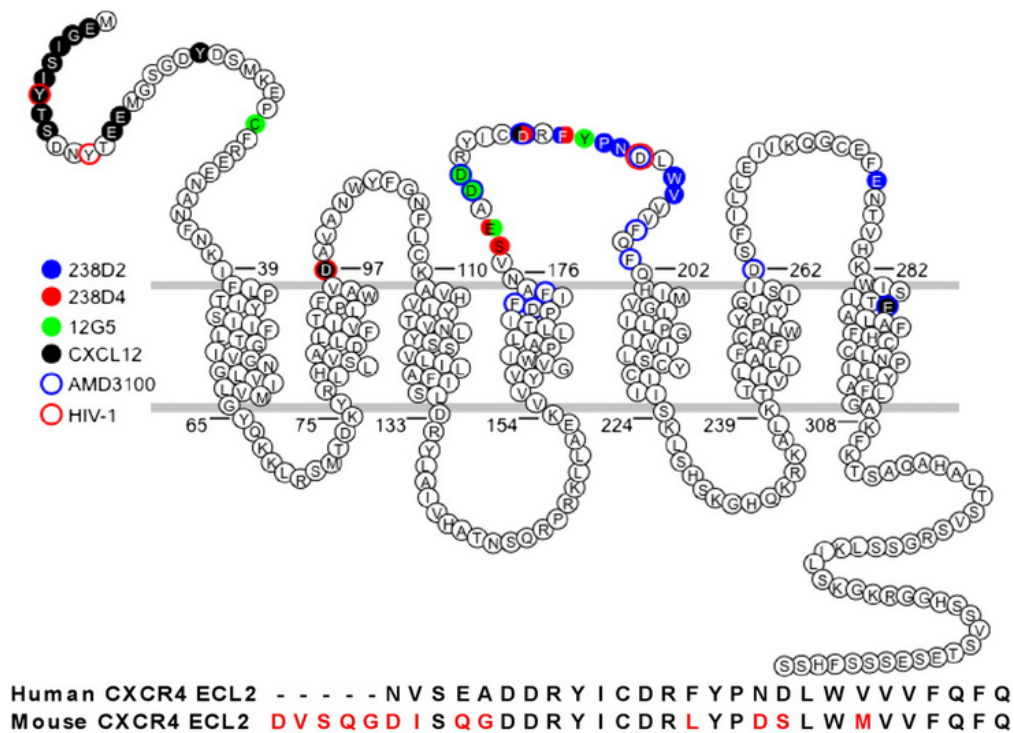

Figure 2.1: CXCR 4 receptor (7 transmembrane domains)

5

2.5-Chemokine ligand 12(CXCL12)/SDF-1α:

SDF-1 is a peptide produced in two forms, SDF-1α/CXCL12a and SDF-1β/CXCL12b, by alternate splicing of the same gene. These are the small cytokine peptide that belong to the intercrine family. The mouse SDF-1α proteins are identical with SDF-1β in the 89 N-terminal amino acids but the beta form has an additional 4 residues at the C-terminus (Nishikawa, Ogawa et al. 1988). The human and mouse predicted protein sequences are approximately 92% identical (Shirozu, Nakano et al. 1995). The calcium responses for SDF-1α–CXCR4 interactions has been studied in CHO cells and other cells a shown in Table 2.4.

(Table 2.4: CXCR4- SDF-1α interactions)

Cell line/type	Techniques used	Conclusion	References
CHO cells Neutrophils Monocytes PBLs	Fura-2 (fluorescence spectrophotometer)	Human stromal cell-derived factor 1 (SDF-1α), acts as the natural ligand for LESTR/fusin (CXCR-4 receptor)	(Oberlin, Amara et al. 1996)
CHO cells PBMCs	Fura-2 (fluorescence spectrophotometer)	SDF-1α induced an increase in intracellular Ca^{2+} and chemotaxis in CXCR-4-transfected cells.	(Bleul, Farzan et al. 1996)

2.6-Chemokine receptors 4(CXCR4): Physiological role

It has been also reported that CXCR4 acts as Ca^{2+} mediated neuromodulator (Guyon, Skrzydelski et al. 2008) and regulates the migration of IEC-6 and CaCo2 cells by release of intracellular Ca^{2+} when stimulated by SDF-1α (dose-dependent manner) (Agle, Vongsa et al. 2010). In rat neonatal cardio-myocytes, SDF-1α acts as cardio-protective chemokine and enhanced the beat frequency (i.e. improved systolic function) by enhancing the Ca^{2+} transient (Hadad, Veithen et al. 2013).

2.7-Data quantification: single cell analysis and heterogeneity

Mostly, the quantification of the cell response to drugs is based on population-average measurements (See Figure 1.3). However the single-cell analysis have demonstrated that there is inherent heterogeneity present in cell populations for many systems (Slack, Martinez et al. 2008). Recent studies suggest that the variability may signify important information that can unravel the classification of cell states (Snijder and Pelkmans 2011). Towards this, a major challenge is to find the biological significance of the cell states (Colman-Lerner,

Gordon et al. 2005). In this context, we need to develop computational tools along with the experimental platform based on high resolution imaging. Characterization of heterogeneity has been carried based on fixed-cell assays/immunofluorescent assays in Hela cells (Slack, Martinez et al. 2008) but limited studies are there to characterize the heterogeneity in live cell dynamic responses. To best of our knowledge, there is no live imaging and analysis of CXCR4 mediated calcium responses in Hela cells.

2.8 Mathematical modeling of calcium oscillations:

Calcium encoding can be mathematically characterized by biophysical models of calcium signaling (Baran 1996, Falcke 2004). Phase plane and bifurcation analysis of can be used to study how the frequency and amplitude of calcium oscillations may vary as a function of any parameter of the model (Giri, Patel et al. 2014). Occurrence of AM-, FM- or AFM-encoding can be assessed on the extension of the minimum to maximum range of amplitude and frequency of Ca^{2+} oscillations and the bifurcation structure of the system under study (Berridge 1997).

(Table 2.5: Mathematical modeling and high dimension (biological) data analysis)

Target data type	Problems related to data	Main conclusion	References
Calcium oscillations	Synchronization of channel cluster opening and closing.	The inter spike interval (ISI) consists of a deterministic and a stochastic part (spontaneous oscillations obeys a Poisson process), calcium oscillations can be described by a time dependent nucleation	(Skupin and Falcke 2007)
Microarray data of metabolite profiles	Integration of different data types	Algorithm proposed for the identification of coordinated patterns; clustering; a parameter that can evaluate the biological significance of the clusters found	(Stegmayer, Gerard et al. 2012)
Gene expression	High dimensional data	Principal component analysis (PCA) as a dimension reduction approach	(Ma and Dai 2011)
Ca^{2+} dynamics	Automated sorting for	Image segmentation to separate highly correlated cells, PCA for dimensional	(Mukamel, Nimmerjah

in large numbers of cells	cells' locations and their dynamics	reduction, Temporal deconvolution and spike detection to extract spike times	n et al. 2009)
Exercise stress test (EST)	High dimensional data with various feature	PCA for dimensional reduction that reduces the training error and the sum of the training and test times.	(Babaoğlu, Fındık et al. 2010)
Biological imaging data	Pattern recognition	Toolbox proposed combining PCA and k-means clustering	(Shamir, Delaney et al. 2010)

2.9-Existing Challenges with respect to cell-to-cell variability

❖ The existing heterogeneity in the single cell responses makes the data analysis rather challenging.

❖ It is difficult to obtain the dose response curve for drug-cell interactions during drug screening.

❖ It is difficult to obtain the specific signature corresponding to a disease (diagnosis) with respect to the cell responses in a large population.

2.10-Objective:

In this study, we plan to develop a generalized framework that can be used for classification of cells with respect to any GPCR mediated calcium response. Such analysis can be used to classify the cells, (on cells, off cells, delayed response, quick response) quantify the cell-to-cell variability in a cell population and dose response relationship.

The purpose of the present work is twofold-

i- This study is aimed at developing an algorithm that allows the classification of the cells in a population with respect to its dynamic properties (during drug-cell interactions).

ii- Validation of the algorithm through construction of dose-response curves for various GPCR targeting drugs such as SDF-1α, endothelin, norepinephrine from real-time imaging of calcium dynamics.

Chapter 3: Materials and methods

3.1-Cell culture

HeLa cells were cultured in MEM medium (Cellgro, Manassas, VA) supplemented with 10% dialyzed fetal bovine serum (Atlanta Biologicals, Flowery Branch, GA) and antibiotics. 0.2×10^6 cells were seeded on 29 mm glass bottom dishes (In Vitro Scientific, Sunnyvale, CA) and maintained in culture until 70–80% confluency (in a 37°C and 5% CO_2 incubator).

3.2-Drug and concentrations:

The time series data for 3 GPCR targeting drugs has been used. Norepinephrine (Sigma, St. Louis, MO) in HBSS was used to activate the Alpha-2 adrenergic receptor at different concentrations (1–200 mM). Stromal derived factor (Sigma, St. Louis, MO) in HBSS was used to activate the CXCR4 receptor at different concentrations (1–800 ng) and Endothelin was used to activate endothelin receptor (50-1000 ng).

3.3-Live cell imaging and image analysis

For the time-series data, Hela cells were imaged on 29 mm glass bottom dishes using Leica-Andor spinning disc confocal imaging system with EM-CCD camera. 20X objective was used to perform calcium imaging of 28 cells/field of view (in Hank's Balanced Salt Solution, HBSS, Invitrogen, Life Technologies, Grand Island, NY) (approximately). Cells were loaded with 2 microM Fluo-4 (Molecular Probes, Life Technologies, Grand Island, NY, Exc: 488nm; Em: 510 nm) for 30 min in HBSS. Then the cells were washed with HBSS for three times (each time 15 min incubation). Time-lapse imaging was performed every second before and after drug addition. Raw image data were analyzed with Andor software to obtain the time course of fluorescence levels in single cells. Image background correction was done by using the Andor background subtraction. For each cell, cytosolic calcium increase was measured by quantifying the fold change of the fluorescence level of Fluo-4 with respect to the basal level. A cell population of size >54 was used for each of the doses of a specific drug.

3.4-Histogram and kernel density analysis:

In order to find the probability distribution of different dynamic properties (peak

intensity of Fluo-4, time required to reach peak intensity and area under the curve), we performed a histogram analysis and kernel density analysis. For comparison of cell population at different drug doses, histograms were plotted and kernel-density function was fitted using MATLAB.

3. 5 Algorithm for the classification of calcium responses:
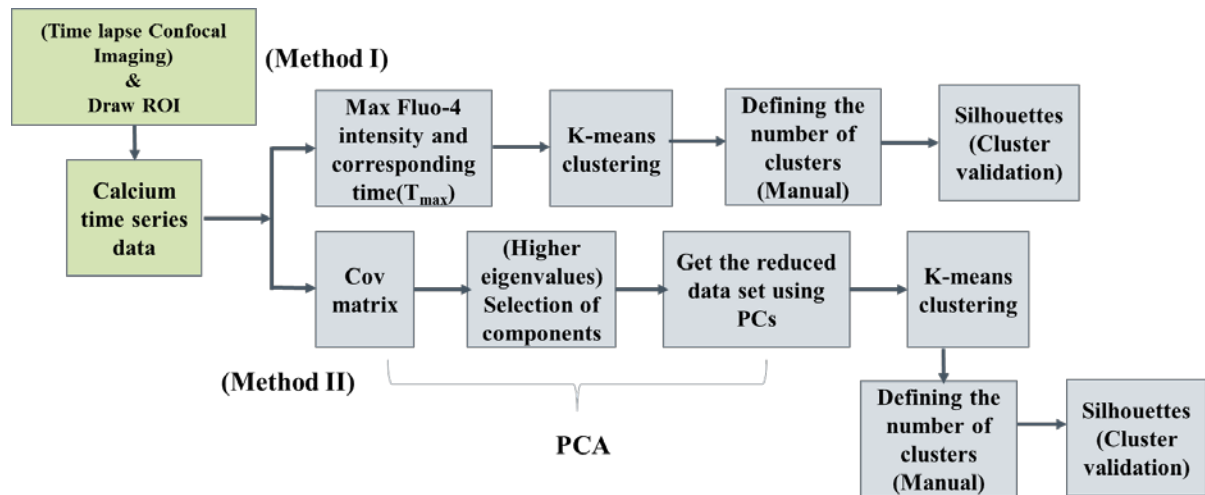
Figure 3.1: Schematic diagram for calcium imaging using microscopy and data analysis

3.6-Feature Extraction using principal component analysis (PCA):

PCA is a dimension reduction technique to project high dimensional data into a new dimensional representation of the data that describes as much of the variance in the data as possible with minimum reconstruction error. Mathematically, PCA is defined as an orthogonal linear transformation that transform the data into new coordinate system such that the greatest variance by any projection of the data comes to lie on the first coordinate (called the first PC), the second greatest variance on the second coordinate, and so on. PCs are calculated using the Eigen value decomposition of a data covariance matrix. The transformation of the dataset to the new principle component axis produces the number of PCs equivalent to the number of original variables (See Figure 3.2). But for many dataset, the first

Figure 3.2: PCA Algorithm

several PCs explain the most of the variances, so the rest can be eliminated with minimal loss of information. PCs whose Eigen values are very small i.e. near to zero are generally eliminated.
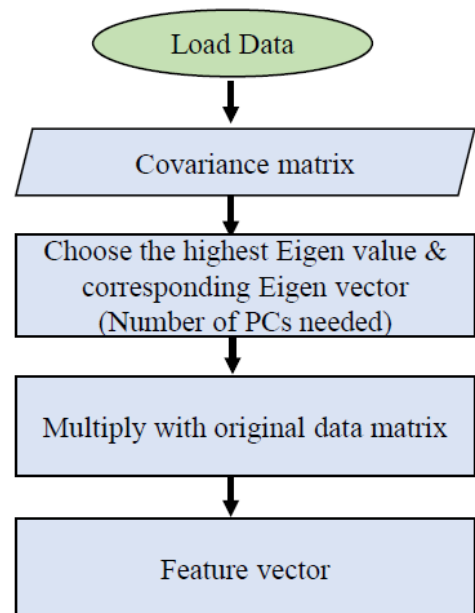
For each cell population, we had a vector of size 56x130 and principal-components analysis (PCA) was used to reduce the dimensionality (time series data for the cell population) from 56x130 to 56x8. The final dimension to be retained was determined through the computation of the eigenvalues of the covariance matrix (Figure 3.3).
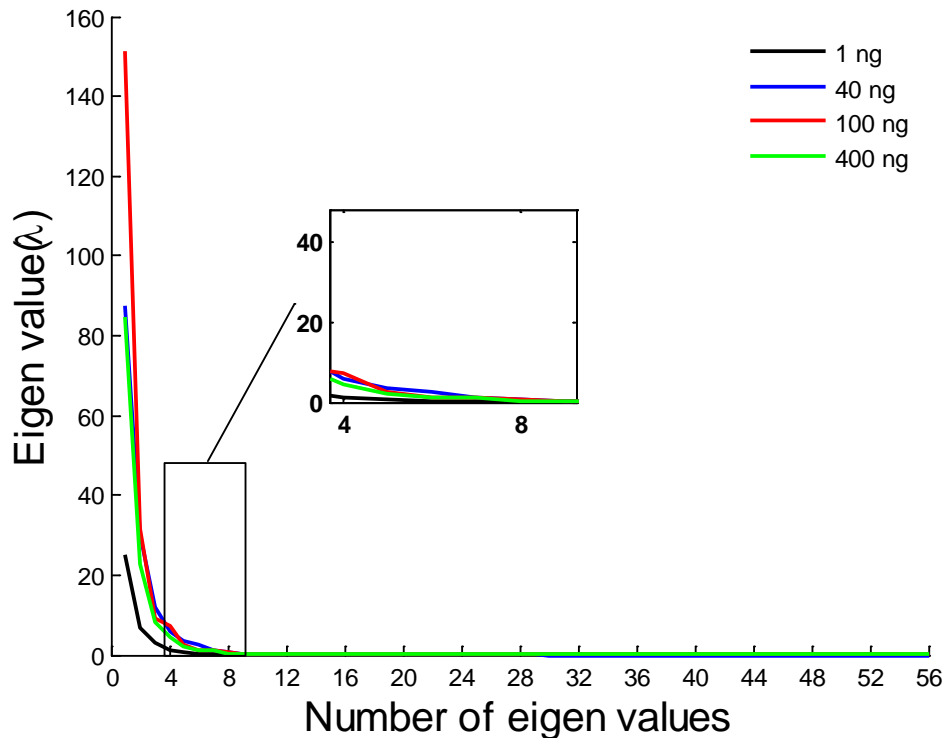
Figure 3.3: Sorted Eigen-value (λ) plot for the data co-variance matrix from Ca^{2+} response imaging data (for various drug doses)

3.6-K-means clustering:

Clustering is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar to each other than to those in other groups. In supervised learning, the computer system is trained using a set of pre-defined classes, and then used to classify unknown objects based on the patterns detected in training. In unsupervised learning there are no classes defined a priori, and the computer system subdivides or clusters the data, usually by using a set of general rules. An example of supervised learning is automatic detection of protein localization, in which the computer system is trained using images of probes for known sub-cellular compartments. An example

11

of unsupervised learning is clustering an expression profiling microarray experiment into groups of genes with similar expression patterns.

Table 3.1: Various K-means

Types of K-means	References
K-means	Khalilian et. Al., 2010
A Hybridized K-means clustering	Dash et al., 2010

The K-means algorithm is one of the unsupervised clustering method. Given a set of numeric objects A and an integer number k, the K-means algorithm searches for a partition of A into k clusters that minimizes the within groups Euclidean distance. The K-means algorithm starts by initializing the k cluster centers.

The steps of the K-means algorithm are written below (Figure 3.4):

1. Initialization: Choose randomly K input vectors (data points) to initialize the clusters.

2. Nearest-neighbor search: For each input vector, find the cluster center that is closest, and assign that input vector to the corresponding cluster.

3. Mean update: update the cluster centers in each cluster using the mean (centroid) of the input vectors assigned to that cluster.

4. Stopping rule: repeat steps 2 and 3 until no more change in the value of the centroids.

K-means clustering was used for the 56x8 vector and the clustered cells were visualized again using principle component analysis (56x2).
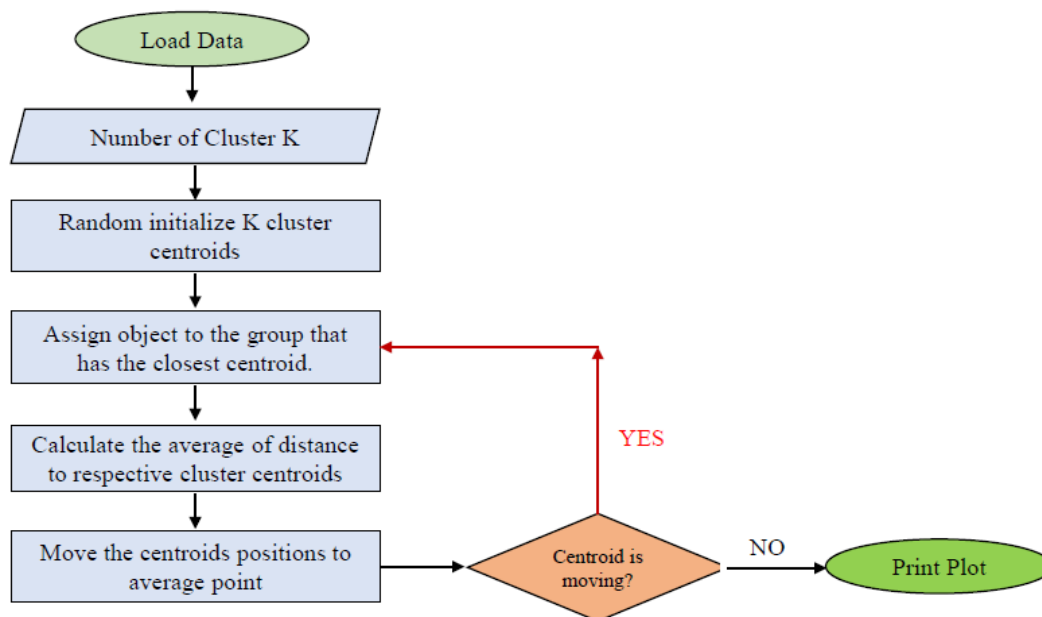

Figure 3.4: K-means clustering algorithm

3.7-Silhouette plot: Cluster validation:

Silhouette values of each point in their own cluster shows how much closely they are related to each other.

Silhouette Value:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i),\ b(i)\}}$$

Which can be written

$$s(i) = \begin{cases} 1 - \dfrac{a(i)}{b(i)}, & if\ a(i) < b(i) \\ 0, & if\ a(i) = b(i) \\ \dfrac{b(i)}{a(i)} - 1, & if\ a(i) > b(i) \end{cases}$$

$a(i)$ is average dissimilarity of i with all other data within the same cluster.

$b(i)$ be the lowest average dissimilarity of i to any other cluster which i is not a member.

3.8-Determination of number of clusters using Silhouette plot:

To determine the number of k in the K-means clustering, we performed simulation of clustering for various k values (k=2, k=3, k=4) (Figure 3.5). For each of the k-values, the performance of the K-means classification is validated by the Silhouette plot. The Silhouette plot for the clustering corresponding to various k's clearly shows that, for k=3, k-means clustering gives the clusters that are tightly grouped (for k=2, the negative Silhouette values indicates the possible presence of another cluster; for k=4, one of the cluster is having very small size: number of cells in the group <2) (Figure 3.5 and 3.6). Hence, using this technique the cell population can be classified with respect to three states for any drug doses.

(Table 3.2: Silhouette values and their significance)

Silhouette value	Conclusion
1.0-0.7	A strong structure has been found
0.7-0.5	A reasonable structure has been found
0.5-0.25	The structure is weak and could be artificial
<0.25	No substantial structure has been found
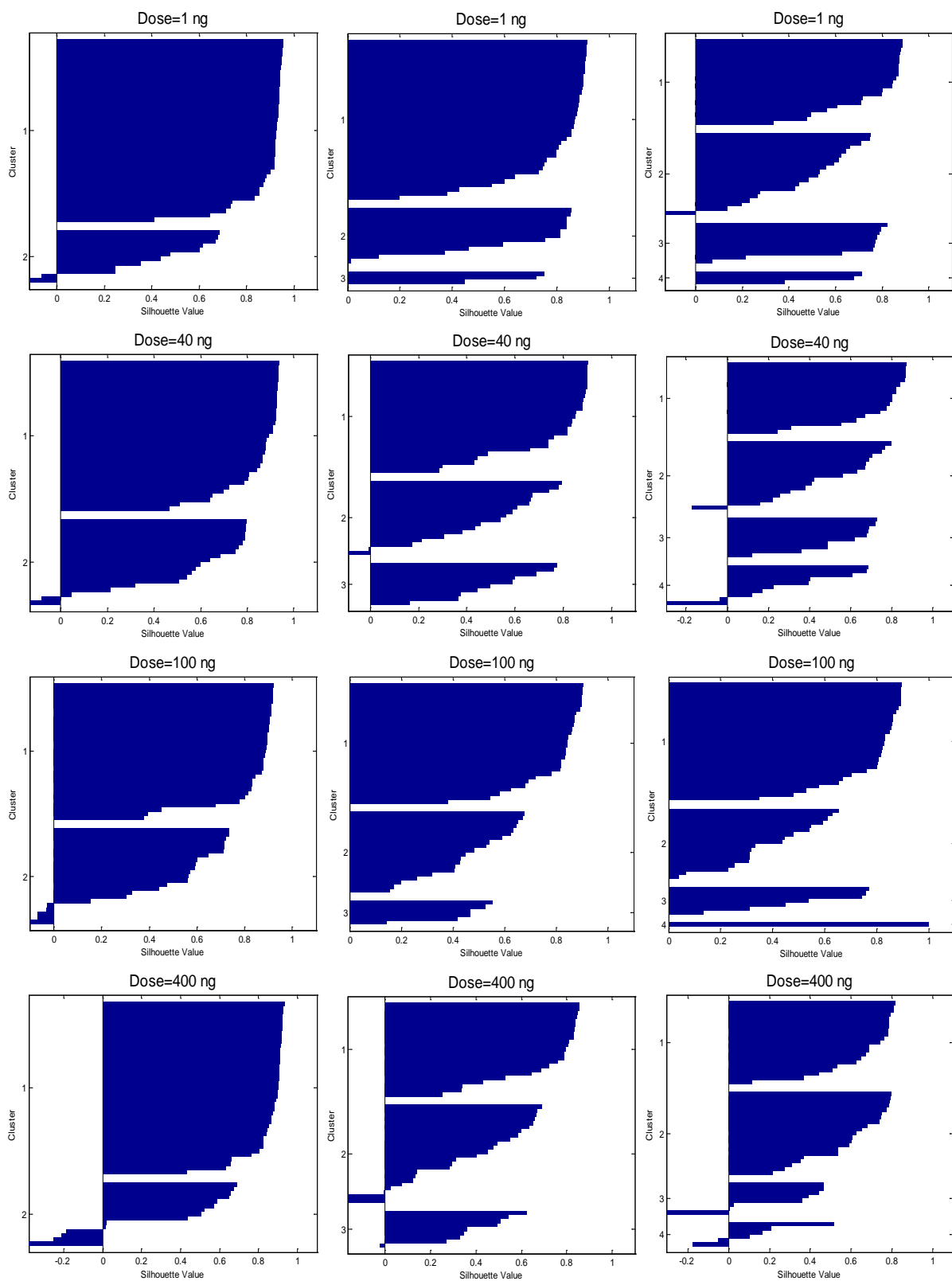
Figure 3.5: Silhouette Clustering (k=2, 3, 4)
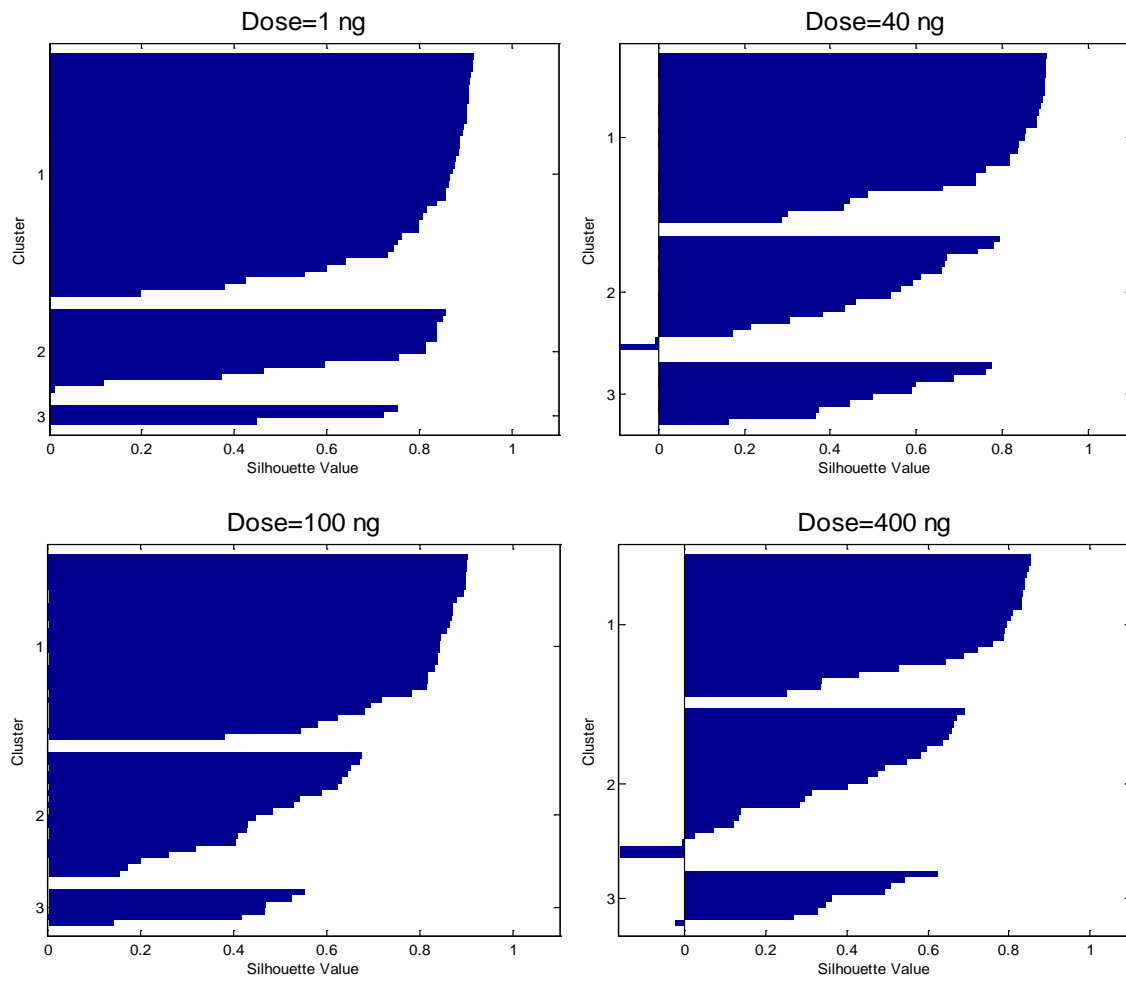
Figure 3.6: Silhouette Clustering for different dose (k=3)

A
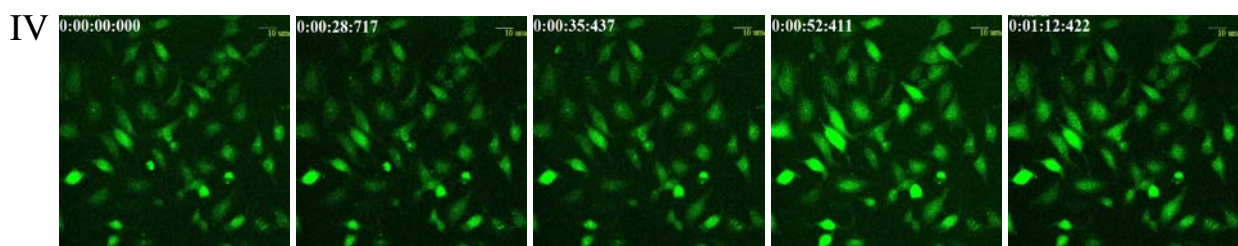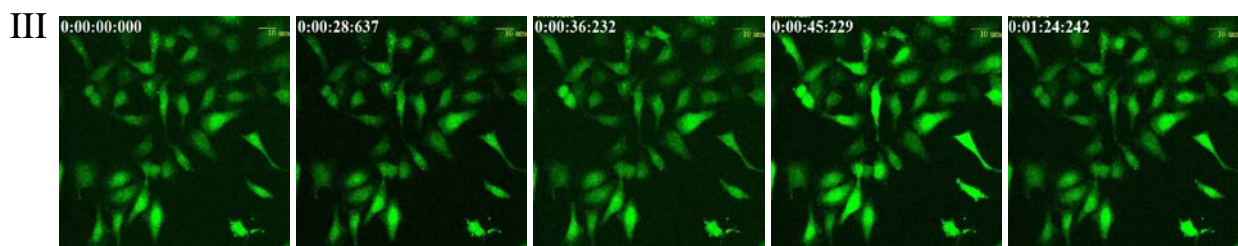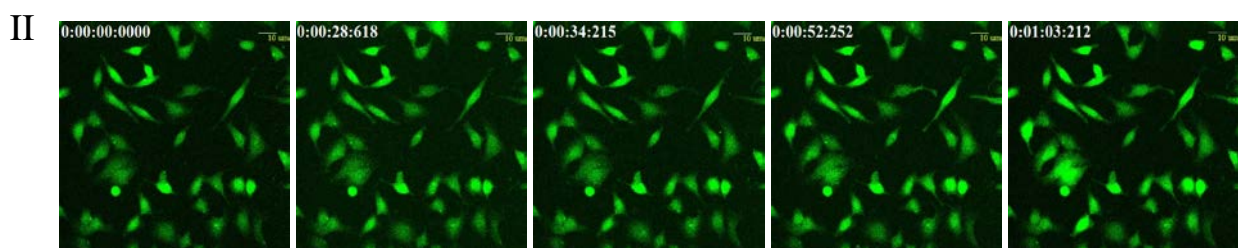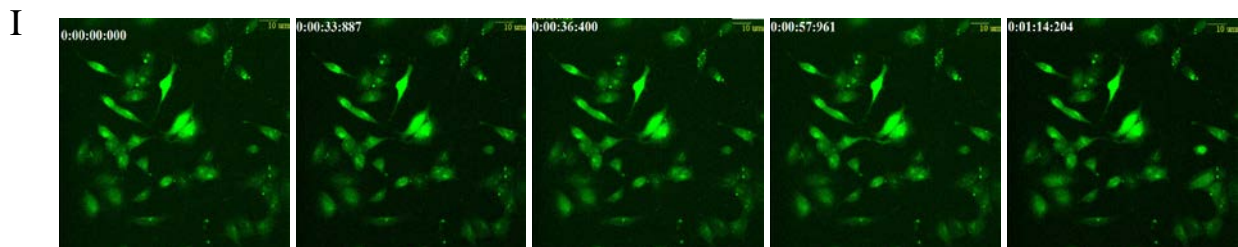
B
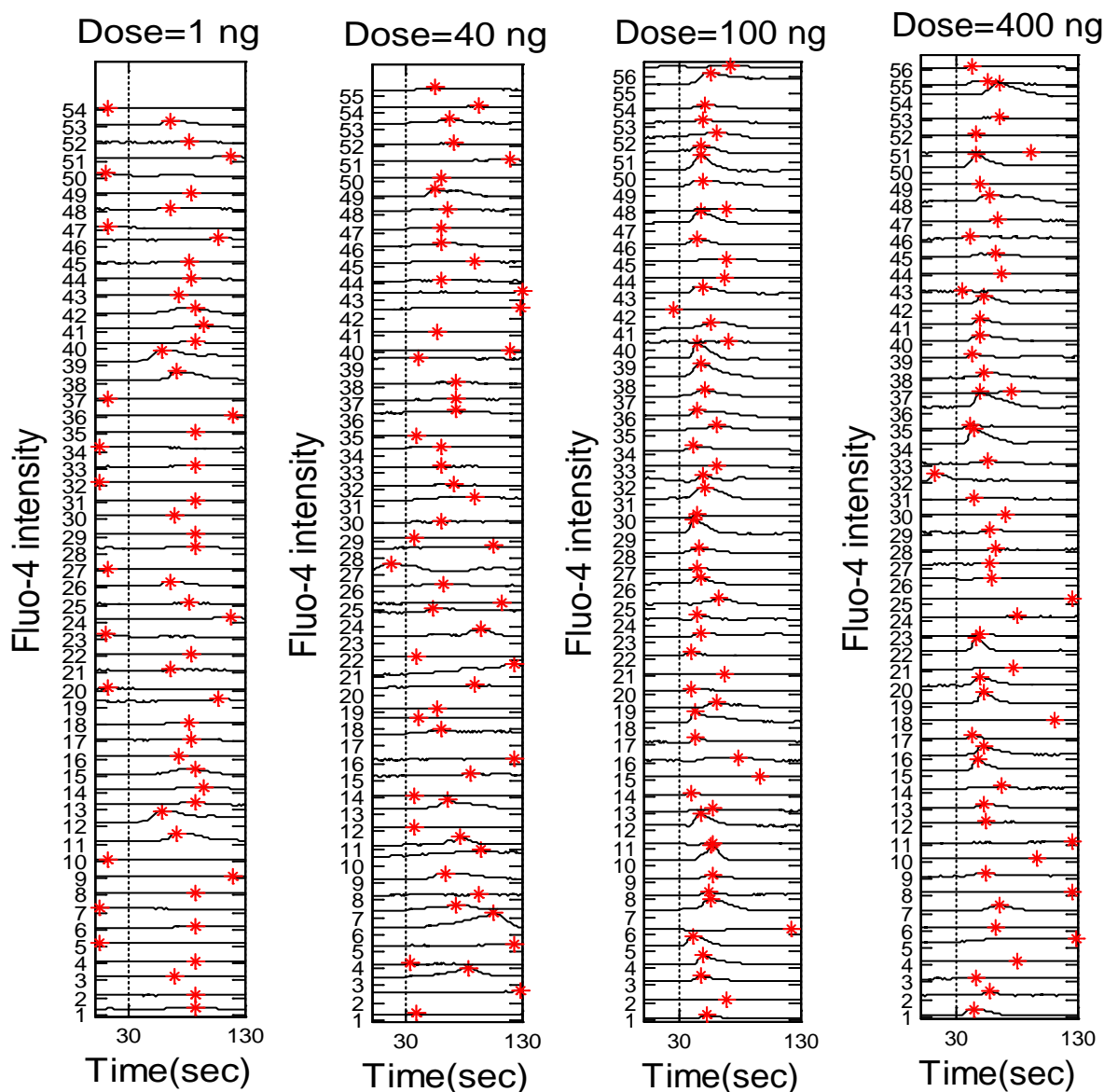
Figure 4.1: Heterogeneity in calcium dynamics of Hela cells during SDF-1α mediated CXCR4 activation (A) Representative images of HeLa Cells showing Ca^{2+} response in presence of SDF-1α (CXCR-4 specific agonist) (I=1 ng, II=40 ng, III=100 ng, III=400 ng). (B) Time course of Ca^{2+} response: SDF-1α was added at 30 s, (*shown by dashed line*) and then response was measured for up to 130s. Maximum Fluo-4 intensity (corresponds to maximum Ca^{2+} response) and time reaching to maximum Fluo-4 intensity (T_{max}) was calculated for each cells (*point highlighted with red star*).

The results clearly shows that the addition the SDF-1α (CXCR-4 specific agonist) induces CXCR4 receptor activation and hence Ca^{2+} responses were observed (Gi-mediated calcium responses). The response was measured by monitoring the Fluo-4 intensity through time lapse imaging. At lower range of drug dose (1 ng and 40 ng) most of the cells were non-responding (note that the maximum Fluo-4 intensity is detected before the drug addition). At low dose very few cells are responding as these cells may have smaller number of the receptors at cell surface (with proper orientation) that can be activated. At a higher dose, (100 ng and 400 ng), most of the cells were showing relatively higher amplitude responses (magnitude of maximum fluorescence).

A

18

B

Probality distribution

C

D

Probality distribution
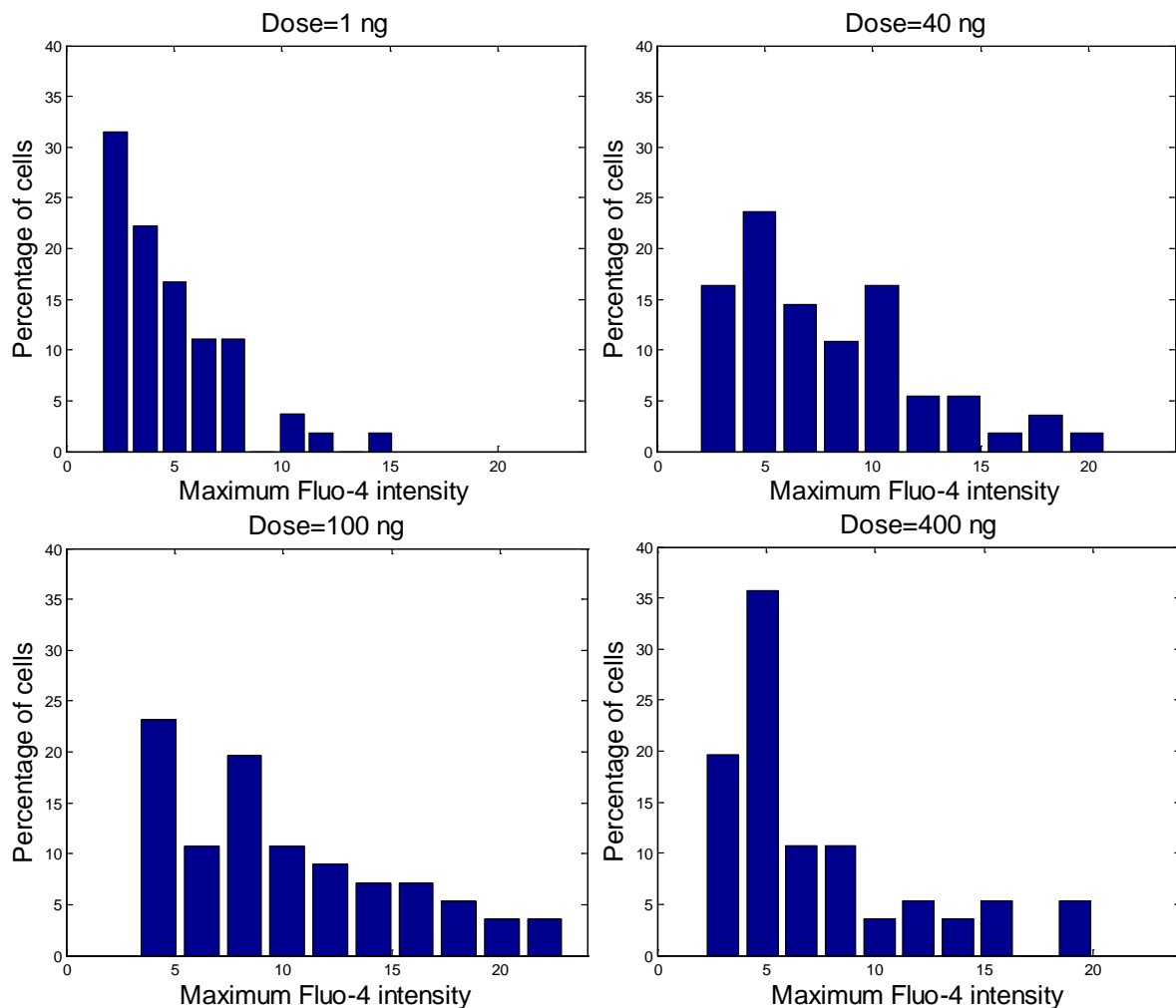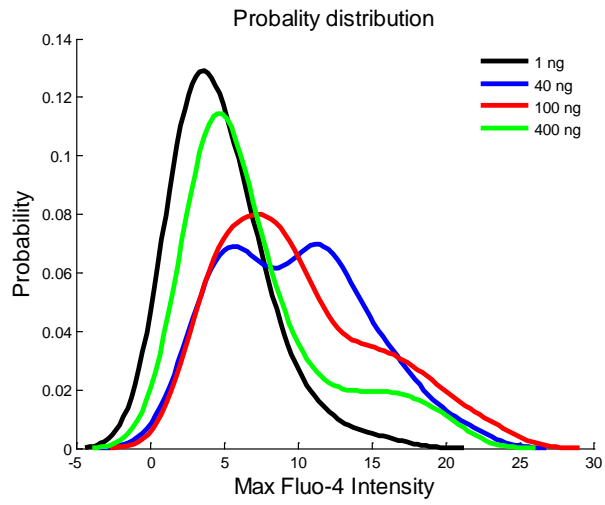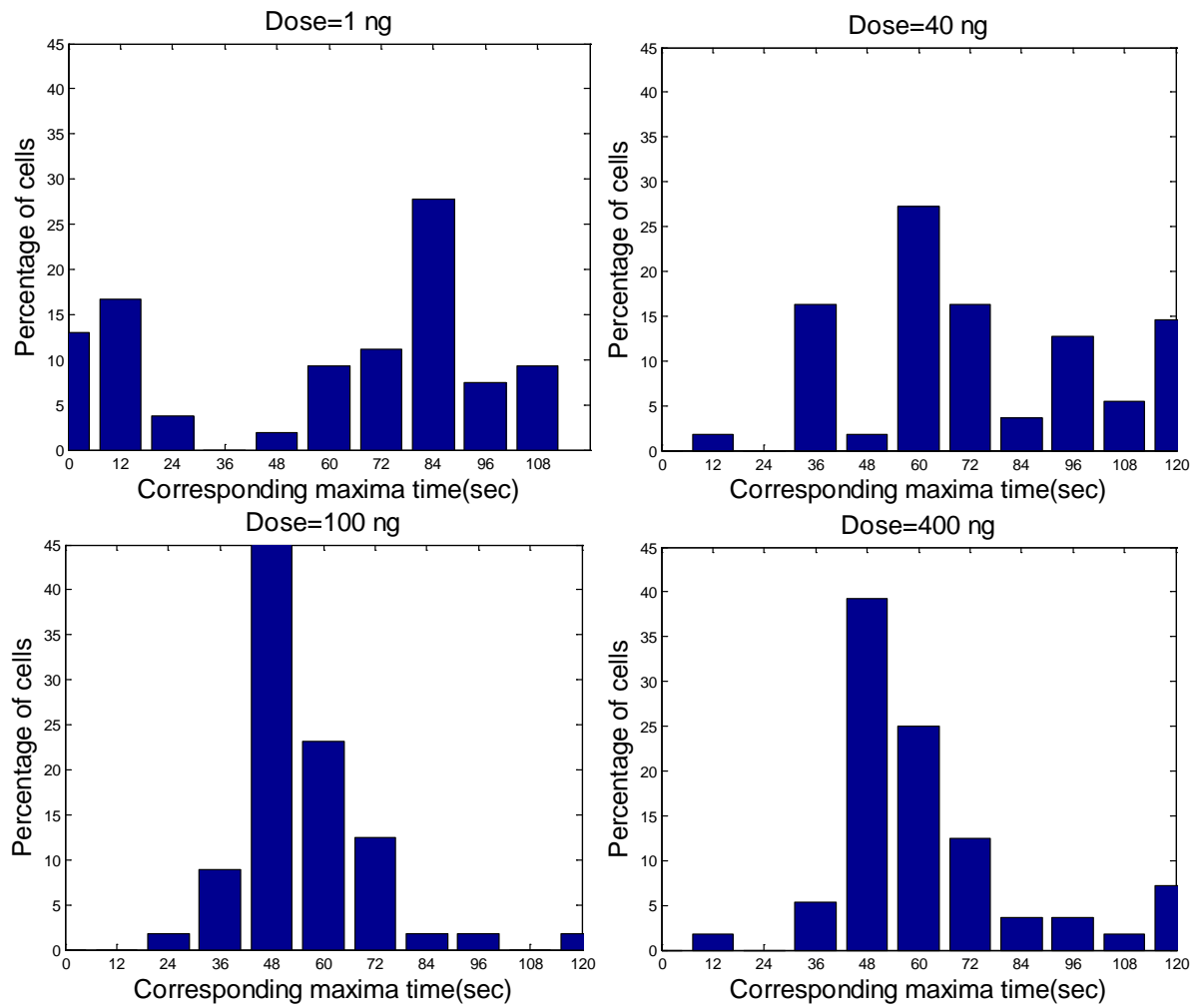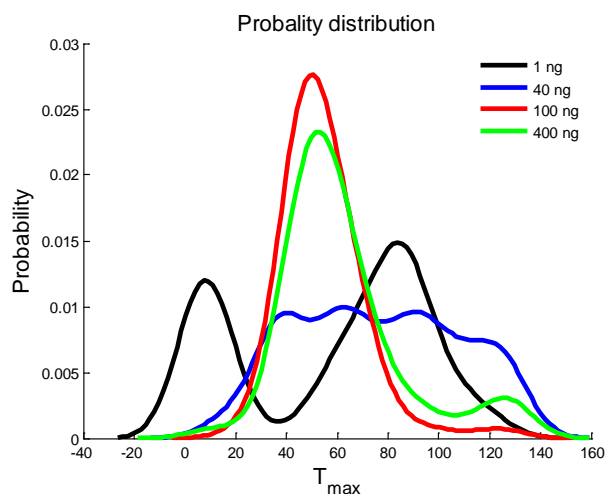
Figure 4.2: Comparison of the probability distribution (cell-to-cell variability) of the calcium response in HeLa cell at various doses of SDF-1α (A) The histogram representation of the maximum Fluo-4 intensity in a cell population at various doses and (B) The kernel density fitted to the maximum Fluo-4 intensity (C_{max}) in a cell population at various doses (C) The histogram representation of the time required to reach maximum Fluo-4 intensity (T_{max}) in a cell population at various doses and (D) The kernel density fitted to the time required to reach maximum Fluo-4 intensity (T_{max}) in a cell population at various doses. The calcium imaging was done at 4 drug doses (1 ng, 40 ng, 100 ng, 400 ng).

Figure 4.2 A and C clearly shows that the variability in the cells significantly changes with change in drug dose. Also the kernel density function fitted to the response characteristics at various drug doses indicates that the probability distribution does not follow a normal distribution (Figure 4.2 B and D). The histogram representation of the maximum Fluo-4 intensity (C_{max}) indicates that with increasing the drug dose, initially, the population mean shifts from a lower C_{max} to higher C_{max} value, and then from higher C_{max} to lower C_{max} value. The respective nonparametric kernel density function also shows that the population mean is toggling between lower and higher C_{max} changes in drug doses. The kernel density fit for T_{max} shows that similar trends exists for T_{max} (T_{max} toggling with drug doses).

A

Dose=1 ng

Dose=40 ng

Dose=100 ng

Dose=400 ng

B

Probality distribution

- 1 ng
- 40 ng
- 100 ng
- 400 ng

Similar toggling effect was noticed even for AUC (as it is in Figure 4.2-B &D). The cell population shifts to a higher AUC with increase in the drug dose followed by a decrease in AUC with further increase in drug dose (toggling).

A

22

Dose=400 ng

B

Dose=1 ng Dose=40 ng Dose=100 ng Dose=400 ng
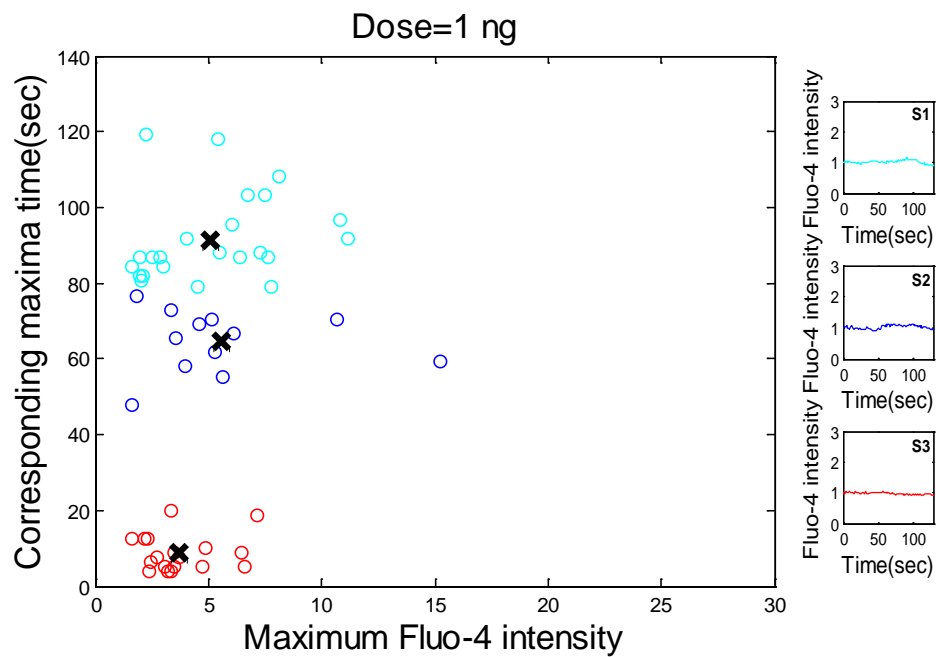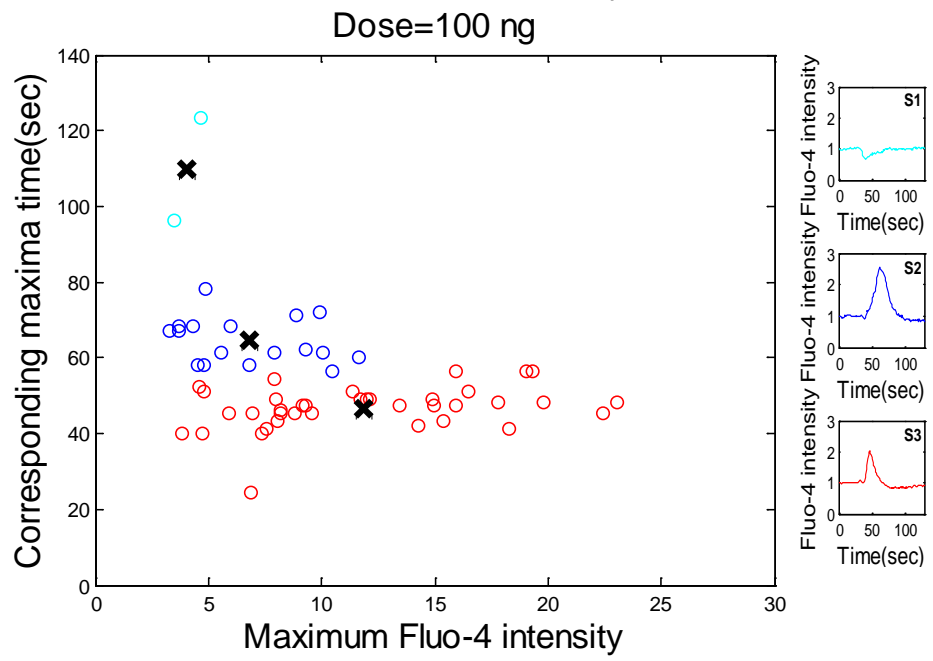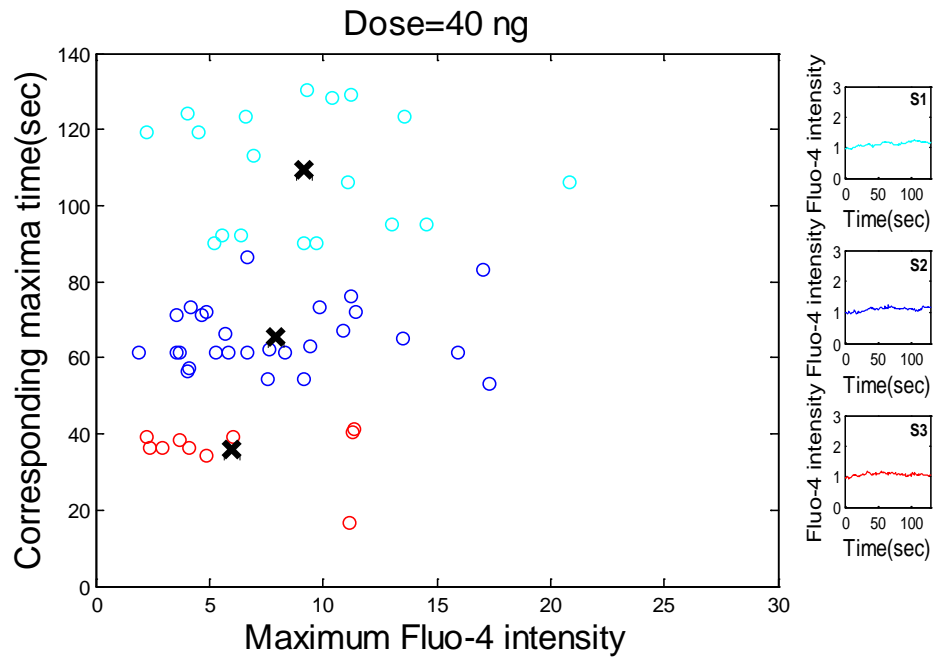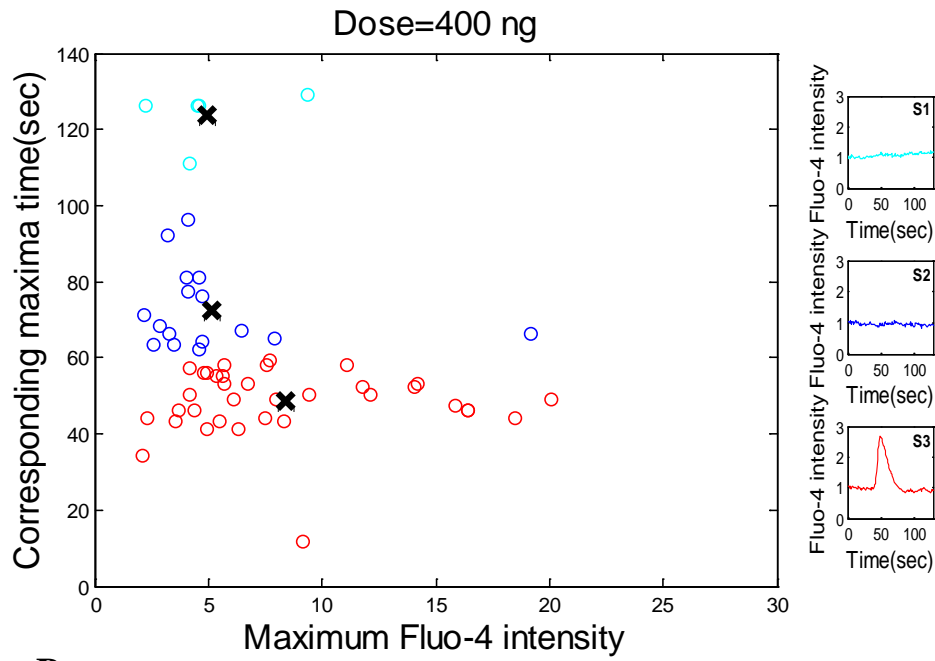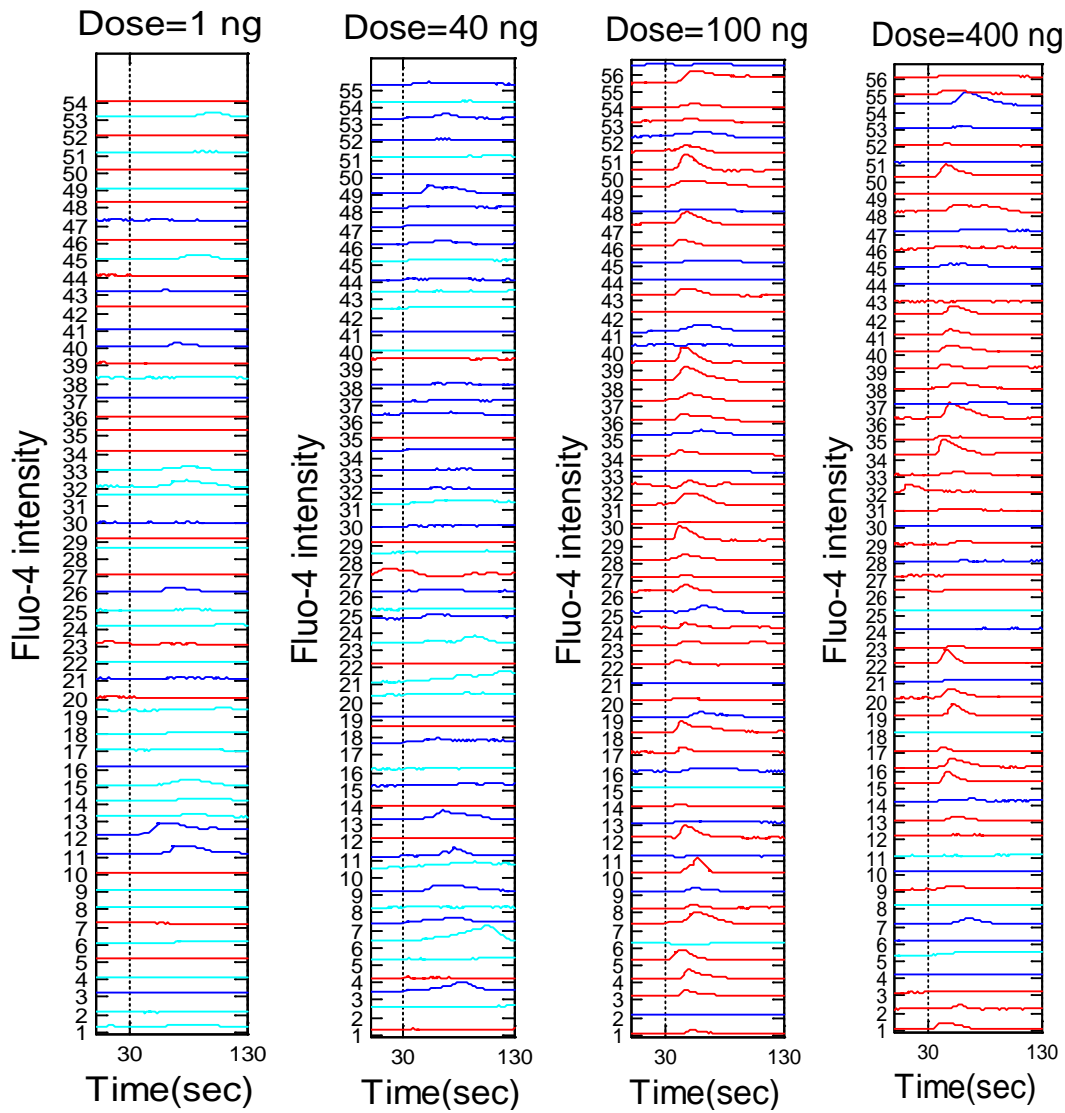
C
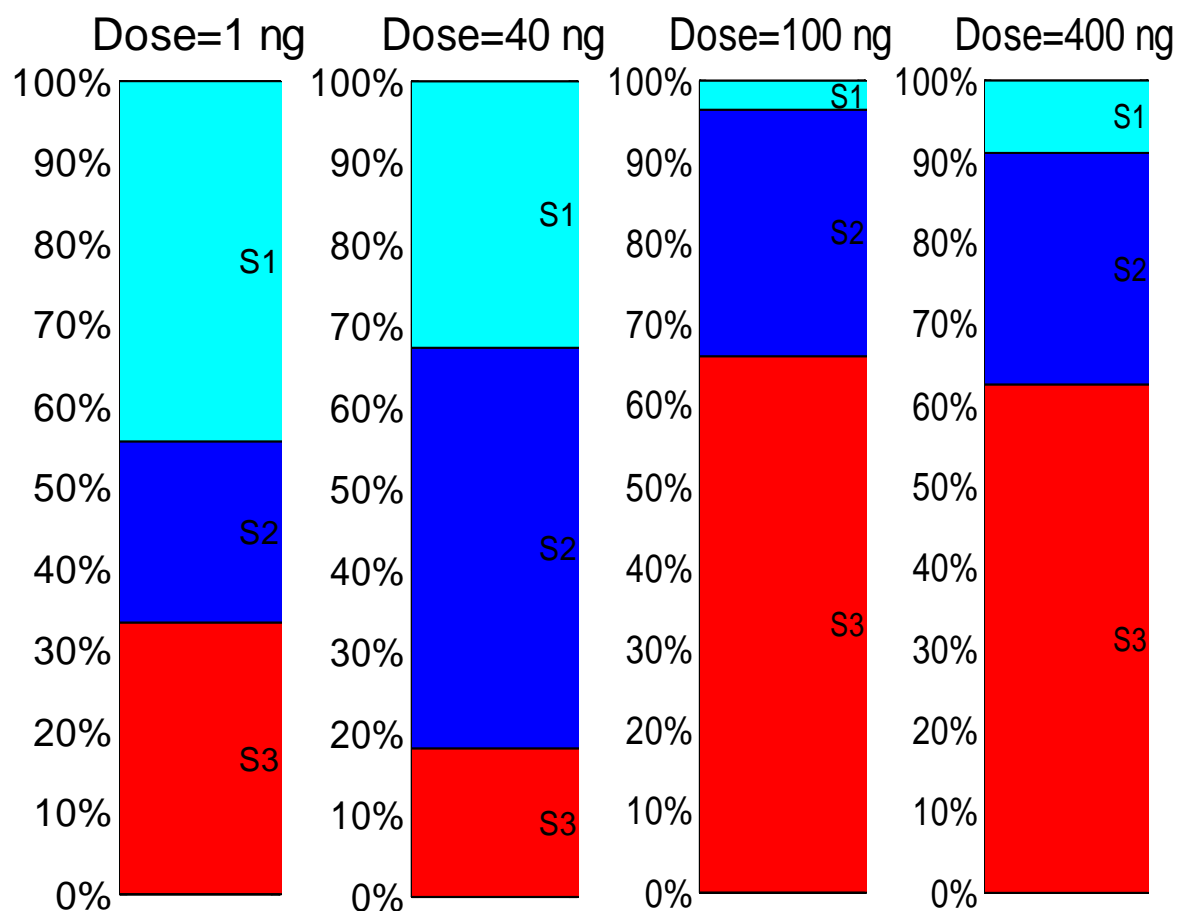
Figure 4.4: Characterization of CXCR4 mediated calcium responses in HeLa cell population as a mixture of subpopulations (method I). (A) Scatter plots: In graph, X-axis is maximum Fluo-4 intensity (corresponds to max Ca^{2+} concentration) and Y-axis is corresponding time required to reach maximum Fluo-4 intensity (T_{max}). Using these two features, K-means ($k=3$) clustering has been performed (representative calcium dynamics plot for each subpopulation is plotted by side- cell response close to the centroid of a cluster). (B) Classification of cell states with respect to Ca^{2+} response at various drug dose **(non-responding, responding-quick, responding-delayed)**. (C) Stacked bar plots representing the relative proportion of each cell state/subpopulation at various drug doses.

We performed K-means classification of the cells in a 2-D plane where we plotted maximum Fluo-4 intensity in X-direction and the time required to reach the maximum (T_{max}) in Y-direction. Considering these two dynamic parameters, k-means (using k=3) clustering was performed and individually three subpopulation was characterized (S1, S2, S3; red, blue

and cyan color respectively). The centroid of each subpopulations were shown to represent the average behavior in each clusters with various drug doses (Figure 4.4a).

At lower drug dose concentration (1 and 40 ng), approximately all the cells were non-responding (as shown in representative plot for each subpopulation with the three colors, Figure 4.4b). On increasing the drug dose concentration (100 ng and 400 ng), most of the cells were responding (as shown in representative plots for each subpopulation, Figure 4.4 a). By comparing the 100 ng and 400 ng drug dose response, 100 ng drug dose yielded the responses in all three subpopulations; indicating it as an optimum dose (Figure 4.4 b). The stacked bar plot for the 100 ng drug dose shows that, almost 95% of the cells were responding. Also, at 100 ng drug dose, the highest proportion of the S3 subpopulation corresponding to the quick responding cells was observed (clearly depicted from stacked bar plot-red subpopulation). At 100 ng drug dose, the high amplitude responsive cells corresponds to S3 subpopulation and lower amplitude cells corresponds to S2 subpopulation.

A

26

Dose=400 ng
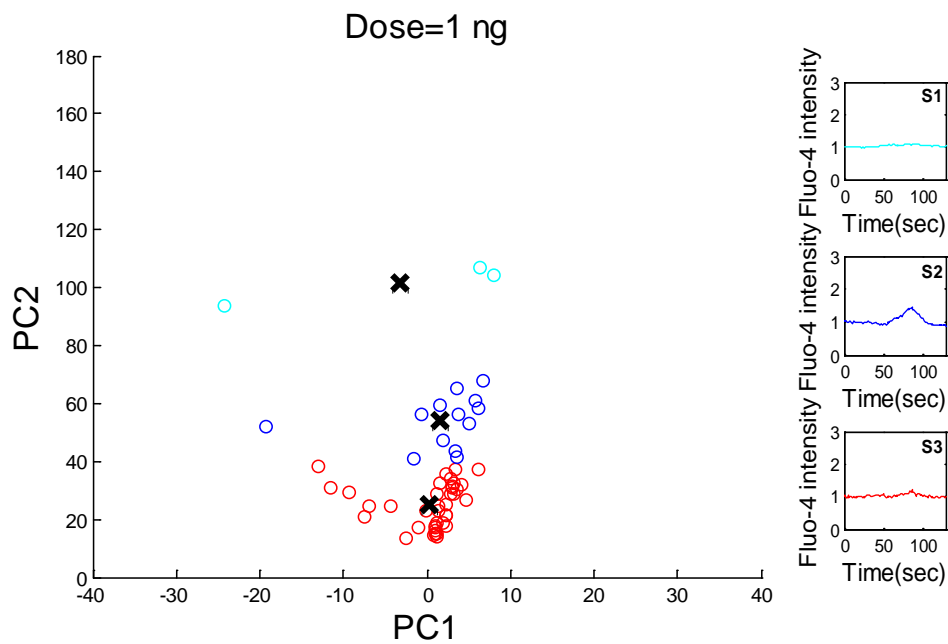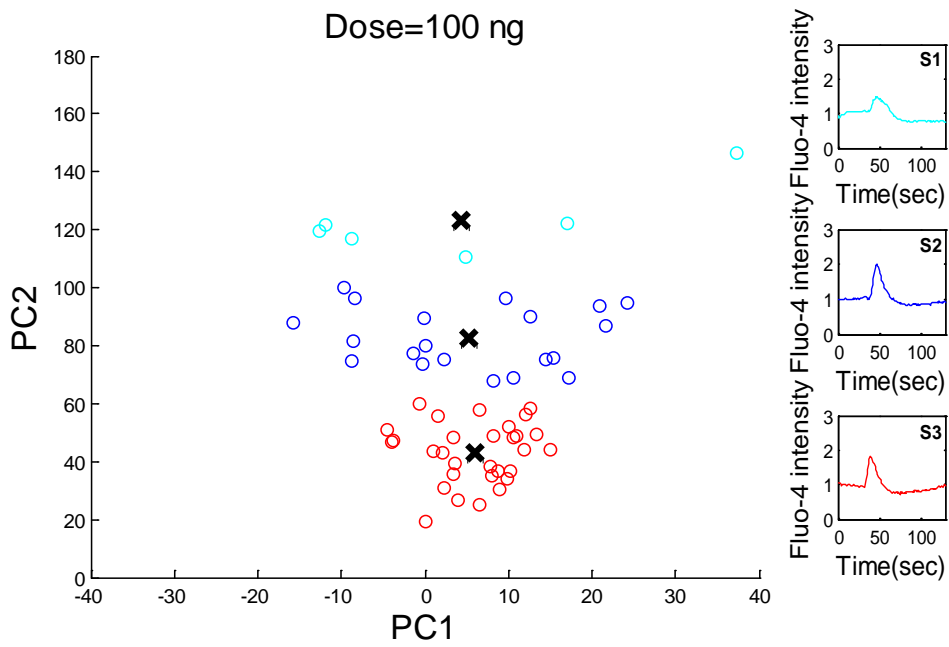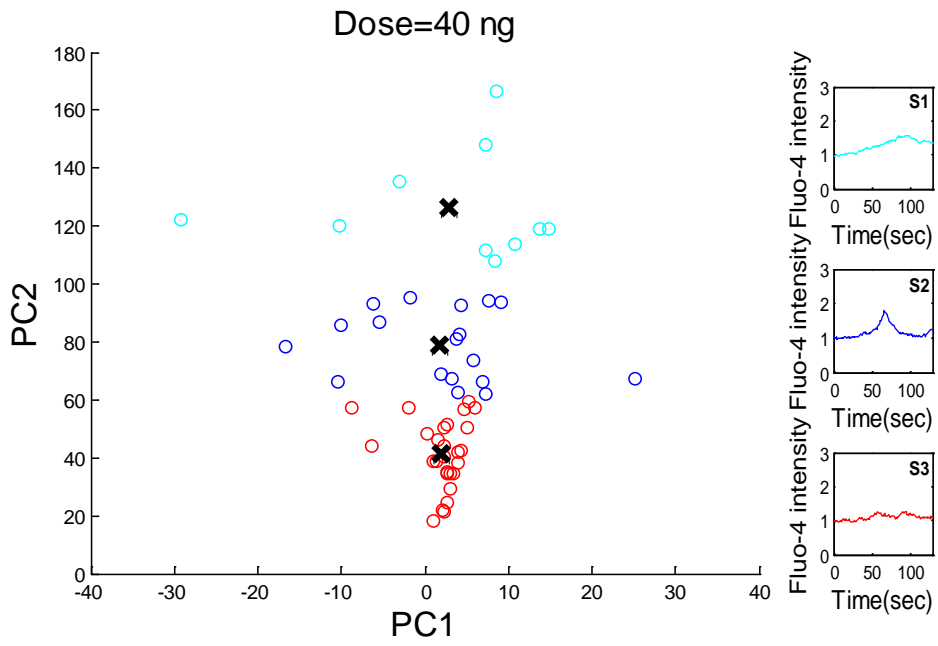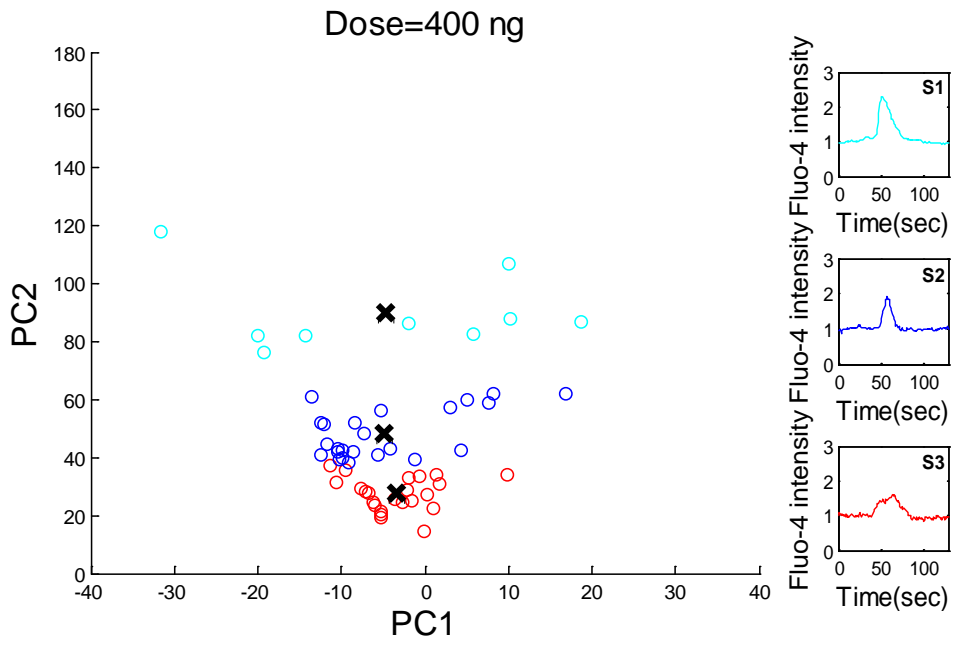
B
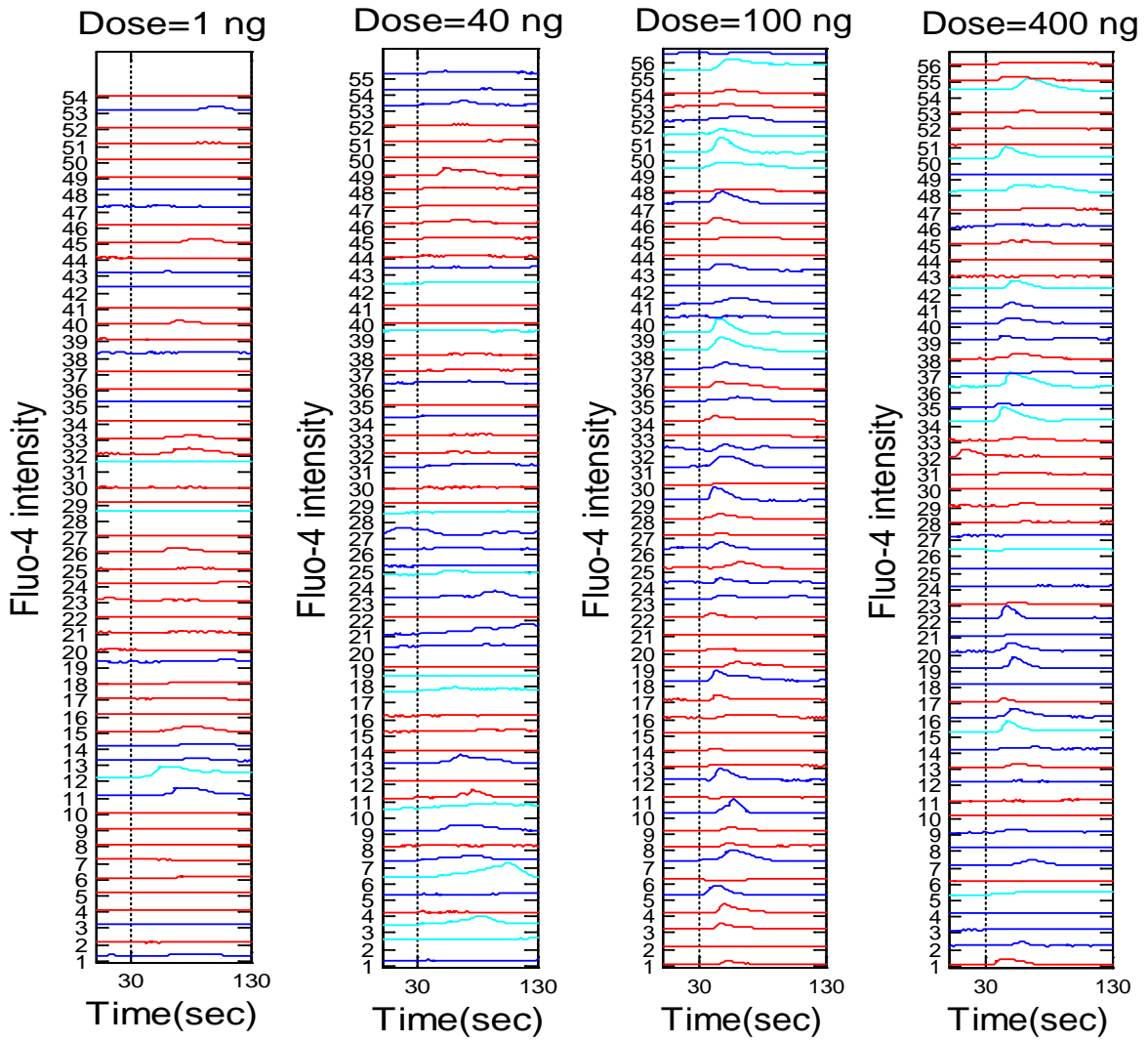
C
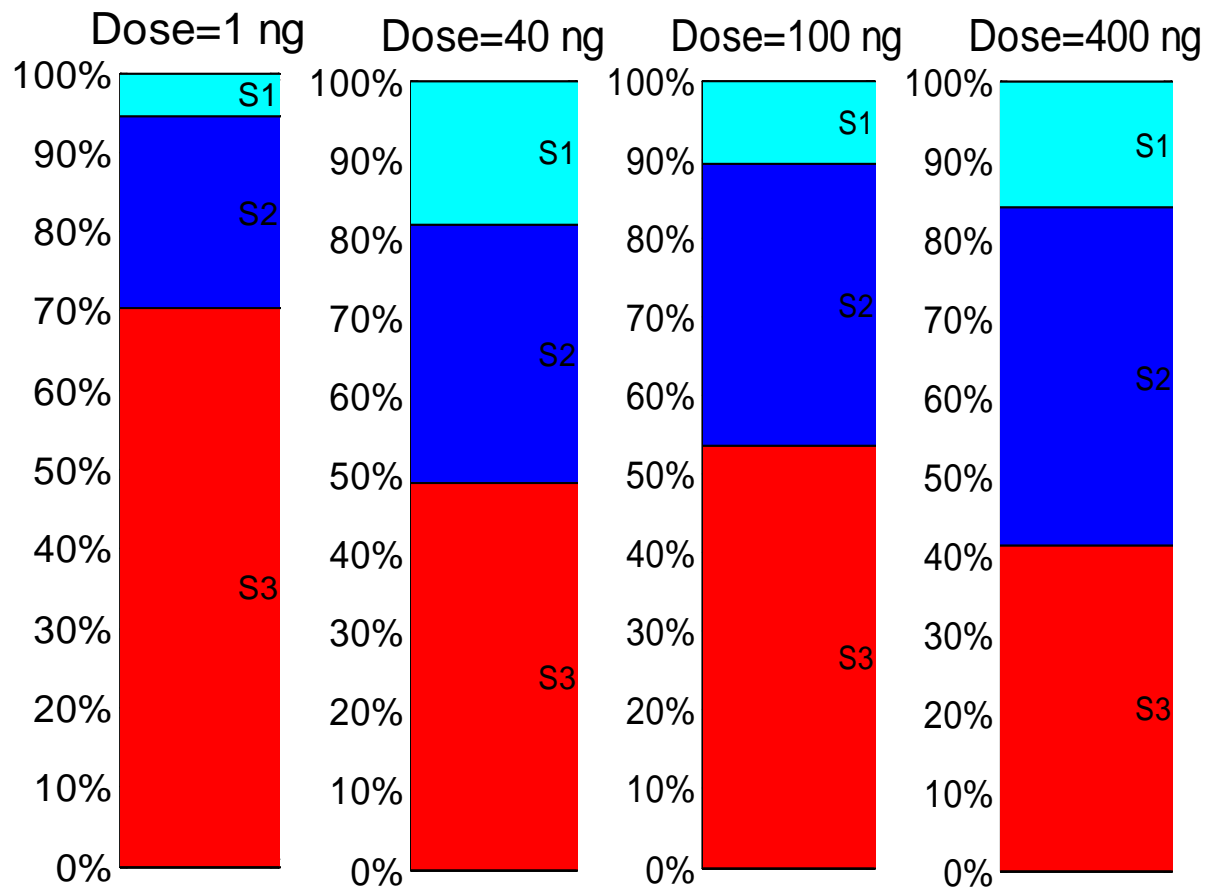
Figure 4.5: Characterization of CXCR4 mediated calcium responses in HeLa cell population as a mixture of subpopulations (method II): Visualization of heterogeneous calcium response in a cell-population as mixture of subpopulation. (A) Method II involves a combination of dimension reduction by PCA and classification using K-means (k-3). (B) Scatter plots: cells are visualized as points using PCA, dimension reduced to two PCs (respective subpopulations are shown in *red, blue and cyan color*). The centroid positions of respective subpopulations are shown in *black cross* (representative response for each of the subpopulation is plotted by side). (B) Classification of cell states with respect to Ca^{2+} response at various drug dose **(low amplitude, high amplitude-quick and high amplitude-delayed).** (C) Stacked bar plots representing the relative proportion of each cell state/subpopulation at various drug doses.

A combination of dimension reduction by PCA (dimension reduced from 130 to 8) and classification using K-means had been used to characterize the cellular heterogeneity in HeLa cells in the presence of SDF-1α (CXCR-4 specific agonist). After classification we further used PCA (from 8 components to 2 components) to visualize the data in two dimension (Figure 4.5 A). The k-means classification results clearly shows that the cells can be classified as multiple cell states (defined as the subpopulations) for each doses. Classification of the cells clearly shows that at low doses (1 ng and 40 ng), most of the cells are non-responsive. Whereas at higher doses (100 ng and 400 ng), most of the cells are responsive with three major states: 1. low amplitude, 2. high amplitude-quick and 3. high amplitude-delayed) (Figure 4.5 B). The stacked bar plot shows that the relative proportion of various cell states as a function of drug doses follow a non-linear trend (Figure 4.5 C). Also it was found that the proportion of the high amplitude-delayed cells is increasing with increasing doses. Such behavior could be due to the presence of negative feedback in the system (inhibition of enzyme through product inhibition or substrate inhibition).

In summary, the cell population clearly demonstrates heterogeneity with varying drug dose concentration. Subpopulation profiles allow more precise quantitative comparison of the responses across doses. As the subpopulation centroid position toggles with drug dose variation, this technique may be implemented for determination for an optimal dose during drug screening.

Similarly for the other GPCR targeting drugs (such as Endothelin and Norepinephrine), the heterogeneity in calcium dynamics were characterized by method-I and method-II (Appendix I, Figure S1,S2, S3 and S4).

Chapter 5: Conclusions

5.1-Status with respect to cell-to-cell heterogeneity

In recent times, biological systems are being investigated at a higher resolution with respect to spatiotemporal dynamics. Such studies reveal the cell-to-cell variability for various cases such as apoptosis, cancer heterogeneity, stem cell differentiation and nuclear factor- κB (Snijder and Pelkmans 2011).   Previous investigations on population-context shows that cell-to-cell variability is present in various systems. For example virus-infection efficiency is heterogeneous and is regulated by cellular crowding and cell-cell-contacts (Snijder, Sacher et al. 2009). However, the study on cell-to-cell variability and clustering of cell states are rather limited in case of drug-cell interaction studies based on live imaging. Specifically, the detailed analysis of cell states on GCPR mediated calcium dynamics has not been performed so far. The main reason behind this is the absence of an integrated platform that includes the imaging, cell segmentation, quantification as well as high-dimensional data analysis. In current study, we propose an integrated scheme for imaging cytosolic calcium in HeLa cells and analyzing the temporal responses in a cell population using clustering techniques. A combination of dimension reduction by PCA and classification using K-means has been successfully used to analyze the cellular heterogeneity and identification of cell states.

5.2-Status with respect to high-dimensional data analysis

Many research groups are taking significant effort in advanced statistical analysis of cell responses (Slack, Martinez et al. 2008, Qiu, Simonds et al. 2011). The heterogeneity of the cancer cell population were characterized as a mixture of phenotypically distinct subpopulations for cancer cells using Gaussian mixture model (Slack, Martinez et al. 2008). They offered a computational framework to assess the cellular heterogeneity and classification of drugs based on heterogeneity. Other groups used approaches based on spanning-tree progression analysis of density-normalized events to analyze the mass spectrometry data from human bone marrow towards analysis of the heterogeneity in multiples parameters from single cells (Qiu, Simonds et al. 2011). Despite recent advancement in the algorithms (Falcke 2004, Skupin and Falcke 2007, De Pittà, Volman et al. 2008, De Pittà, Volman et al. 2009, Mukamel, Nimmerjahn et al. 2009, Dupont, Combettes et al. 2011, Honarnejad, Kirsch et al. 2013), it remains challenging to analyze high-dimensional cell-response data, as the techniques to be selected on the basis of the

biological hypothesis. Also it will depend on the characteristics of specific system of interest. Here we attempt to create a platform for the analysis of cell-drug interactions when calcium dynamics is measured in live cells.

5.3 Major findings

We found that the dynamic properties of the calcium responses shows significant variation when they are stimulated using the GPCR targeting drugs. Here we implemented K-means based clustering to classify the cells with respect to their dynamic properties. We found that approximately two cell states are possible at the lower dose and three cell states are possible at the higher doses. The high dimension data (for large cell number and large number of time points) can be visualized using principle component analysis. We demonstrate that our method is robust to different types of GPCR-based drugs.

5.4 Application of the technique

The clissiication technique proposed can be used for identification of distinct patterns corresponding to drug cell interactions from time-series data obtained through live cell imaging. It can also be used for other systems in order to compare cell populations with pharmacological treatments (different drugs and doses) and genetic perturbations (protein overexpression and SiRNA knockdown studies).

5.5 Advantages of the technique:

The proposed algorithm enables automated analysis and extraction of feature from high-dimension data. Additionally, we were able to perform an unbiased analysis on all the arrays. The Feature extraction in 2-D can be used for formulation of a predictive framework using interpolation.

5.6 Disadvantages of the technique:

The number of groups in a cell population (subpopulations) needs to be fixed by the analyzer before analysis i.e. the algorithm itself cannot choose the number of clusters. The classification across all doses are not on the basis of similar classes/same principal components. The proposed algorithm does not help in explaining the biophysical mechanism.

Chapter 6: Future Scope

6.1-Improvement on the computational technique:

The existing technique can be further improved by clubbing the data for all the doses and performing PCA to find principal components common for all doses. Using those components we can find the reduced matrix for all the doses. In addition to K-means clustering we may investigate other clustering techniques such as SVM/GMM and their efficiency in classification. Additionally, further analysis can be done using other statistical techniques such as box plot, hypothesis testing (ANOVA) and correlation techniques. In order to predict the responses for a given drug dose, interpolation can be used with the 2-d plots of the high-dimension data.

6.2-Proposed experimental Project:

Study of cell cycle phases in stem cells using the proposed technique and immunological staining for specific proteins
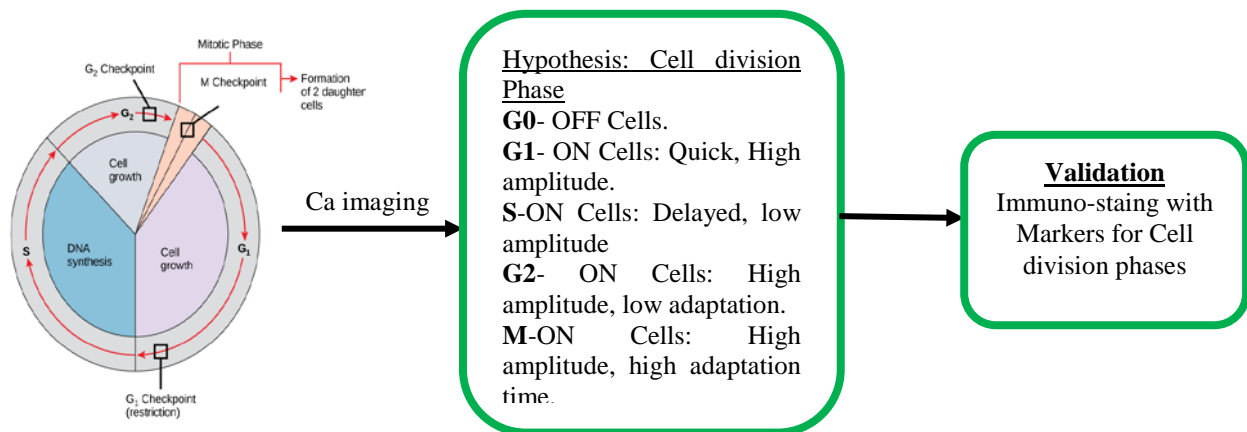
Reference

Aggarwal, B. B., et al. (2007). "Targeting cell signaling pathways for drug discovery: An old lock needs a new key." Journal of Cellular Biochemistry **102**(3): 580-592.

Agle, K. A., et al. (2010). "Calcium mobilization triggered by the chemokine CXCL12 regulates migration in wounded intestinal epithelial monolayers." J Biol Chem **285**(21): 16066-16075.

Babaoğlu, I., et al. (2010). "Effects of principle component analysis on assessment of coronary artery diseases using support vector machine." Expert Systems with Applications **37**(3): 2182-2185.

Badea, T., et al. (2001). "Calcium imaging of epileptiform events with single-cell resolution." Journal of Neurobiology **48**(3): 215-227.

Baran, I. (1996). "Calcium and cell cycle progression: possible effects of external perturbations on cell proliferation." Biophys J **70**(3): 1198-1213.

Berridge, M. J. (1997). "The AM and FM of calcium signalling." Nature **386**(6627): 759-760.

Berridge, M. J., et al. (1998). "Calcium - a life and death signal." Nature **395**(6703): 645-648.

Berridge, M. J., et al. (2000). "The versatility and universality of calcium signalling." Nat Rev Mol Cell Biol **1**(1): 11-21.

Bleul, C. C., et al. (1996). "The lymphocyte chemoattractant SDF-1 is a ligand for LESTR/fusin and blocks HIV-1 entry." Nature **382**(6594): 829-833.

Busillo, J. M. and J. L. Benovic (2007). "Regulation of CXCR4 signaling." Biochimica et Biophysica Acta (BBA) - Biomembranes **1768**(4): 952-963.

Choi, W. T., et al. (2012). "Drug discovery research targeting the CXC chemokine receptor 4 (CXCR4)." J Med Chem **55**(3): 977-994.

Clapham, D. E. (2007). "Calcium Signaling." Cell **131**(6): 1047-1058.

Colman-Lerner, A., et al. (2005). "Regulated cell-to-cell variation in a cell-fate decision system." Nature **437**(7059): 699-706.

De Pittà, M., et al. (2009). "Multimodal encoding in a simplified model of intracellular calcium signaling." Cognitive Processing **10**(1): 55-70.

De Pittà, M., et al. (2008). "Coexistence of amplitude and frequency modulations in intracellular calcium dynamics." Physical Review E **77**(3): 030903.

Dupont, G., et al. (2011). "Calcium Oscillations." Cold Spring Harbor Perspectives in Biology **3**(3).

Falcke, M. (2004). "Reading the patterns in living cells —the physics of ca2+ signaling." Advances in Physics **53**(3): 255-440.

Gao, X., et al. (2014). "Galectin-3 induces cell migration via a calcium-sensitive MAPK/ERK1/2 pathway." Oncotarget **5**(8): 2077-2084.

Gee, K. R., et al. (2000). "Chemical and physiological characterization of fluo-4 Ca(2+)-indicator dyes." Cell Calcium **27**(2): 97-106.

Giri, L., et al. (2014). "A G-Protein Subunit Translocation Embedded Network Motif Underlies GPCR Regulation of Calcium Oscillations." Biophysical Journal **107**(1): 242-254.

Guo, F., et al. (2015). "CXCL12/CXCR4: a symbiotic bridge linking cancer cells and their stromal neighbors in oncogenic communication networks." Oncogene.

Guyon, A., et al. (2008). "Stromal-cell-derived factor 1α/CXCL12 modulates high-threshold calcium currents in rat substantia nigra." European Journal of Neuroscience **28**(5): 862-870.

Hadad, I., et al. (2013). "Stroma Cell-Derived Factor-1α Signaling Enhances Calcium Transients and Beating Frequency in Rat Neonatal Cardiomyocytes." PLoS ONE **8**(2): e56007.

Honarnejad, K., et al. (2013). "Development and Implementation of a High-Throughput Compound Screening Assay for Targeting Disrupted ER Calcium Homeostasis in Alzheimer's Disease." PLoS ONE **8**(11): e80645.

Honarnejad, K., et al. (2013). "FRET-Based Calcium Imaging: A Tool for High-Throughput/Content Phenotypic Drug Screening in Alzheimer Disease." <u>Journal of Biomolecular Screening</u> **18**(10): 1309-1320.

Hutchins, B. I., et al. (2013). "Calcium release-dependent actin flow in the leading process mediates axophilic migration." <u>J Neurosci</u> **33**(28): 11361-11371.

Iino, M. (2010). "Spatiotemporal dynamics of Ca(2+) signaling and its physiological roles." <u>Proceedings of the Japan Academy. Series B, Physical and Biological Sciences</u> **86**(3): 244-256.

Ivannikov, M., et al. (2013). "Synaptic Vesicle Exocytosis in Hippocampal Synaptosomes Correlates Directly with Total Mitochondrial Volume." <u>Journal of Molecular Neuroscience</u> **49**(1): 223-230.

Ivannikov, Maxim V. and Gregory T. Macleod "Mitochondrial Free Ca2+ Levels and Their Effects on Energy Metabolism in Drosophila Motor Nerve Terminals." <u>Biophys J</u> **104**(11): 2353-2361.

Jähnichen, S., et al. (2010). "CXCR4 nanobodies (VHH-based single variable domains) potently inhibit chemotaxis and HIV-1 replication and mobilize stem cells." <u>Proceedings of the National Academy of Sciences</u> **107**(47): 20565-20570.

Ma, S. and Y. Dai (2011). "Principal component analysis based methods in bioinformatics studies." <u>Brief Bioinform</u> **12**(6): 714-722.

Michelini, E., et al. (2010). "Cell-based assays: fuelling drug discovery." <u>Anal Bioanal Chem</u> **398**(1): 227-238.

Mukamel, E. A., et al. (2009). "Automated Analysis of Cellular Signals from Large-Scale Calcium Imaging Data." <u>Neuron</u> **63**(6): 747-760.

Nishikawa, S., et al. (1988). "B lymphopoiesis on stromal cell clone: stromal cell clones acting on different stages of B cell differentiation." <u>Eur J Immunol</u> **18**(11): 1767-1771.

Oberlin, E., et al. (1996). "The CXC chemokine SDF-1 is the ligand for LESTR/fusin and prevents infection by T-cell-line-adapted HIV-1." <u>Nature</u> **382**(6594): 833-835.

Poenie, M., et al. (1986). "Calcium rises abruptly and briefly throughout the cell at the onset of anaphase." Science **233**(4766): 886-889.

Qiu, P., et al. (2011). "Extracting a cellular hierarchy from high-dimensional cytometry data with SPADE." Nat Biotech **29**(10): 886-891.

Shamir, L., et al. (2010). "Pattern Recognition Software and Techniques for Biological Image Analysis." PLoS Computational Biology **6**(11): e1000974.

Shirozu, M., et al. (1995). "Structure and chromosomal localization of the human stromal cell-derived factor 1 (SDF1) gene." Genomics **28**(3): 495-500.

Skupin, A. and M. Falcke (2007). "Statistical properties and information content of calcium oscillations." Genome Inform **18**: 44-53.

Slack, M. D., et al. (2008). "Characterizing heterogeneous cellular responses to perturbations." Proc Natl Acad Sci U S A **105**(49): 19306-19311.

Snijder, B. and L. Pelkmans (2011). "Origins of regulated cell-to-cell variability." Nat Rev Mol Cell Biol **12**(2): 119-125.

Snijder, B., et al. (2009). "Population context determines cell-to-cell variability in endocytosis and virus infection." Nature **461**(7263): 520-523.

Stegmayer, G., et al. (2012). "Data Mining Over Biological Datasets: An Integrated Approach Based on Computational Intelligence." Computational Intelligence Magazine, IEEE **7**(4): 22-34.

Tsai, F.-C., et al. (2014). "A polarized Ca2+, diacylglycerol and STIM1 signalling system regulates directed cell migration." Nat Cell Biol **16**(2): 133-144.

Vachel, L., et al. (2015). "The low PLC-δ1 expression in cystic fibrosis bronchial epithelial cells induces upregulation of TRPV6 channel activity." Cell Calcium **57**(1): 38-48.

Wegner, S. A., et al. (1998). "Genomic organization and functional characterization of the chemokine receptor CXCR4, a major entry co-receptor for human immunodeficiency virus type 1." J Biol Chem **273**(8): 4754-4760.

Wu, B., et al. (2010). "Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists." Science **330**(6007): 1066-1071.

Zlotnik, A. and O. Yoshie (2012). "The Chemokine Superfamily Revisited." Immunity **36**(5): 705-716.

Appendix I: Other GPCR and drugs drug interactions

CXCR4-SDF-1α)(n=28)
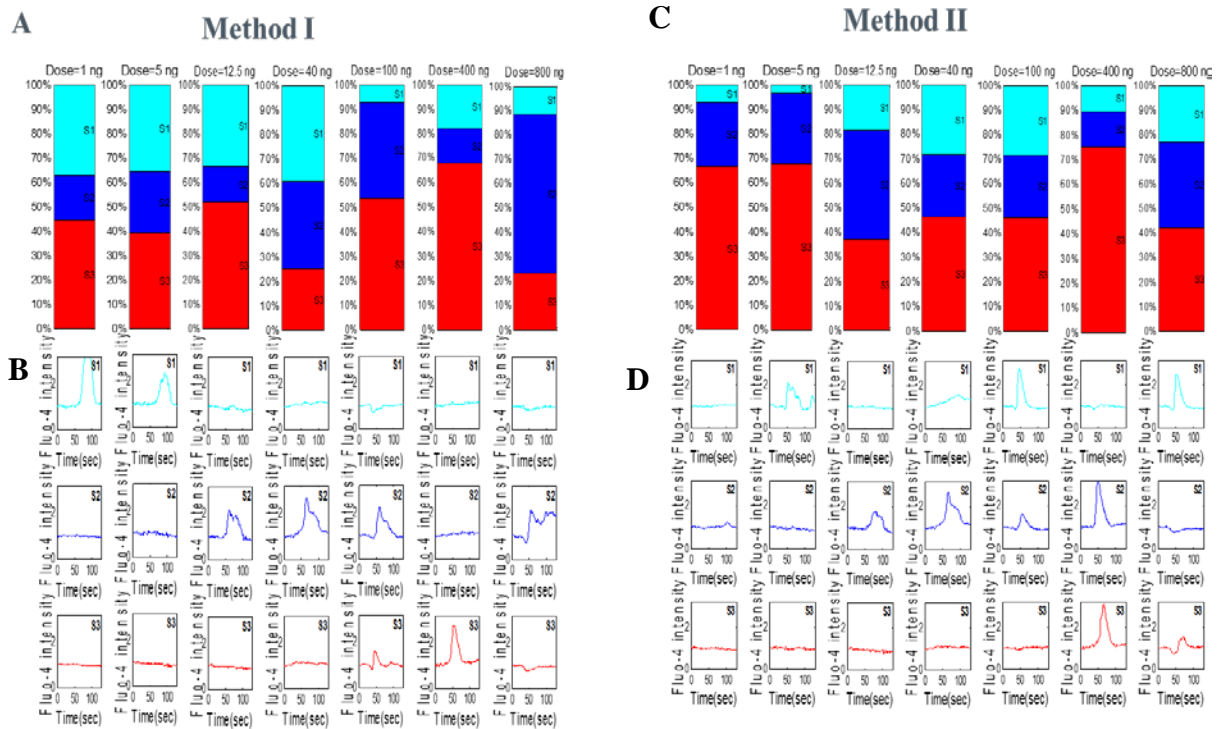
Figure S1: Characterization of cell-to-cell variability as a mixture of subpopulations for CXCR4-SDF-1α (A) using stacked bar plots (method I) (B) classified subpopulation representative (C) Characterized cells subpopulations using stacked bar plots (method II). (D) classified subpopulation representative (method II).

ET-R-Endothelin
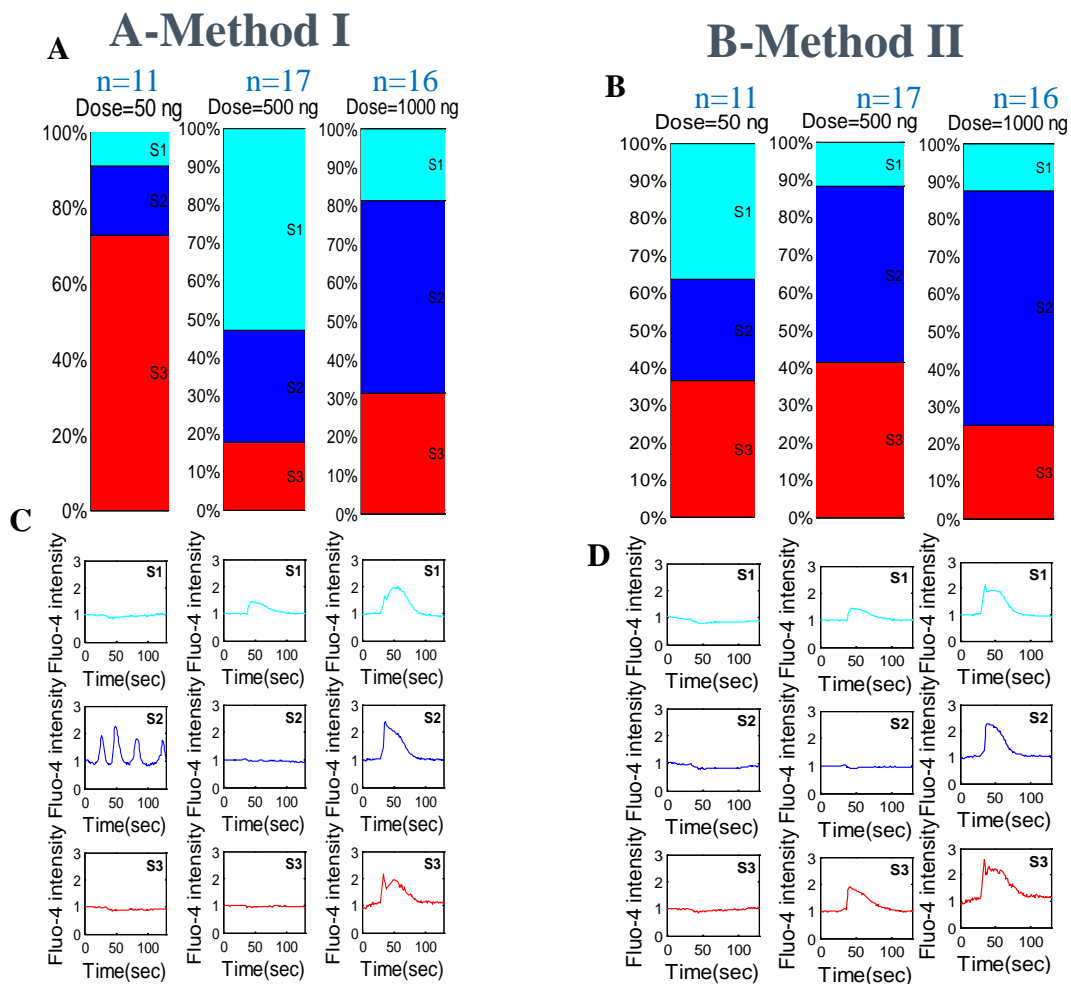
Figure S2: Characterization of cell-to-cell variability as a mixture of subpopulations for ETR-endotheline (A) using stacked bar plots (method I) (B) classified subpopulation representative (C) Characterized cells subpopulations using stacked bar plots (method II). (D) classified subpopulation representative (method II).

α2-Adrenergic receptor-Norepinephrine (Drug dose variation)
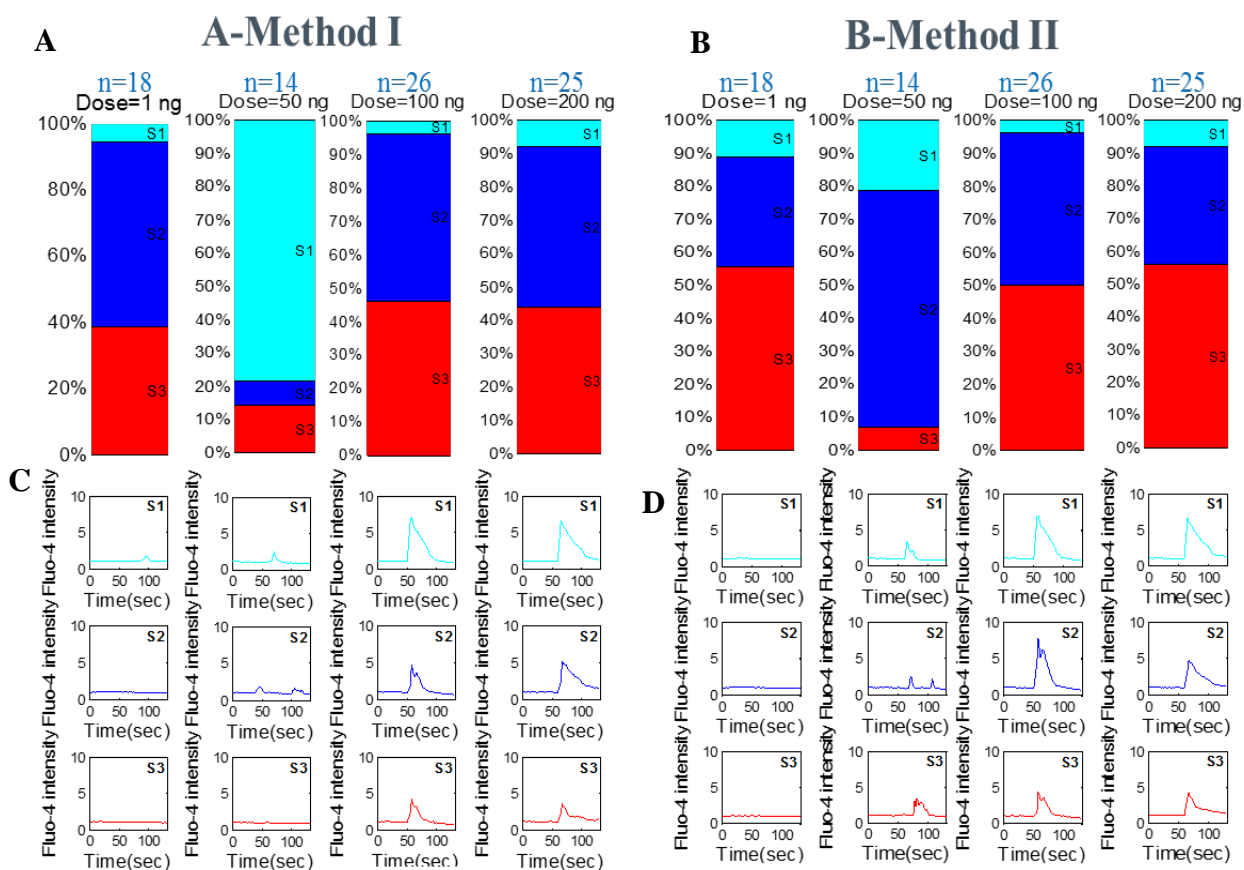
Figure S3: Characterization of cell-to-cell variability as a mixture of subpopulations for α2-Adrenergic receptor-Norepinephrine (drug dose variations) (A) using stacked bar plots (method I) (B) classified subpopulation representative (C) Characterized cells subpopulations using stacked bar plots (method II). (D) classified subpopulation representative (method II).

α2-Adrenergic receptor-Norepinephrine (Cell Number variation)
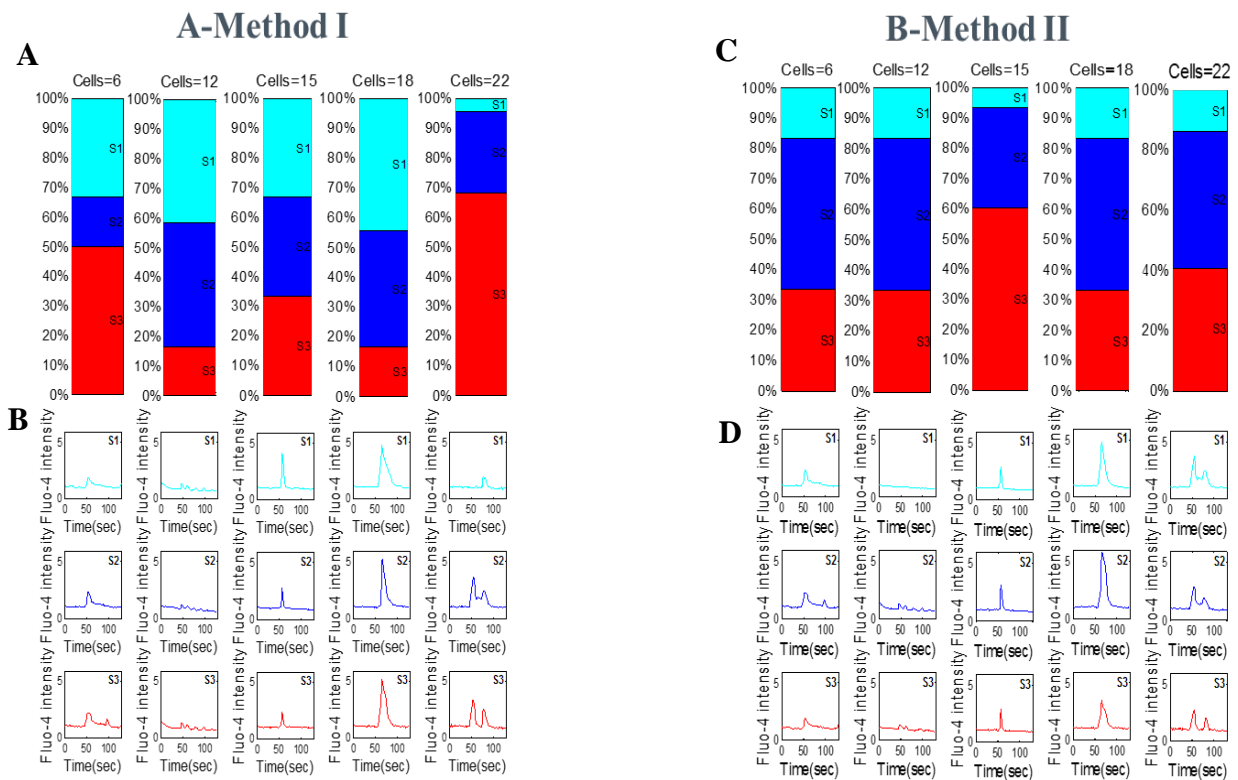
Figure S4: Characterization of cell-to-cell variability as a mixture of subpopulations for α2-Adrenergic receptor-Norepinephrine (cell number variations) (A) using stacked bar plots (method I) (B) classified subpopulation representative (C) Characterized cells subpopulations using stacked bar plots (method II). (D) classified subpopulation representative (method II).

Appendix II: Amino acid sequences of GPCRs and drug peptids

CXCR4- amino acid sequences

```
10          20          30          40          50
MEGISIYTSD  NYTEEMGSGD  YDSMKEPCFR  EENANFNKIF  LPTIYSIIFL
        60          70          80          90         100
TGIVGNGLVI  LVMGYQKKLR  SMTDKYRLHL  SVADLLFVIT  LPFWAVDAVA
       110         120         130         140         150
NWYFGNFLCK  AVHVIYTVNL  YSSVLILAFI  SLDRYLAIVH  ATNSQRPRKL
       160         170         180         190         200
LAEKVVYVGV  WIPALLLTIP  DFIFANVSEA  DDRYICDRFY  PNDLWVVVFQ
       210         220         230         240         250
FQHIMVGLIL  PGIVILSCYC  IIISKLSHSK  GHQKRKALKT  TVILILAFFA
       260         270         280         290         300
CWLPYYIGIS  IDSFILLEII  KQGCEFENTV  HKWISITEAL  AFFHCCLNPI
       310         320         330         340         350
LYAFLGAKFK  TSAQHALTSV  SRGSSLKILS  KGKRGGHSSV  STESESSSFH


SS
```

CXCL12- amino acid sequences

```
10          20          30          40          50
MNAKVVVVLV  LVLTALCLSD  GKPVSLSYRC  PCRFFESHVA  RANVKHLKIL
```

```
              60         70         80         90
NTPNCALQIV ARLKNNNRQV CIDPKLKWIQ EYLEKALNKR FKM
```

α2-Adrenergic receptor
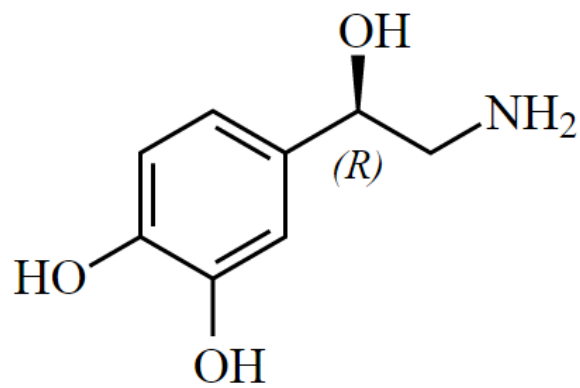
Norepinephrine

```
       10         20         30         40         50
MGSLQPDAGN ASWNGTEAPG GGARATPYSL QVTLTLVCLA GLLMLLTVFG
       60         70         80         90        100
NVLVIIAVFT SRALKAPQNL FLVSLASADI LVATLVIPFS LANEVMGYWY
      110        120        130        140        150
FGKAWCEIYL ALDVLFCTSS IVHLCAISLD RYWSITQAIE YNLKRTPRRI
      160        170        180        190        200
KAIIITVWVI SAVISFPPLI SIEKKGGGGG PQPAEPRCEI NDQKWYVISS
      210        220        230        240        250
CIGSFFAPCL IMILVYVRIY QIAKRRTRVP PSRRGPDAVA APPGGTERRP
      260        270        280        290        300
NGLGPERSAG PGGAEAEPLP TQLNGAPGEP APAGPRDTDA LDLEESSSSD
      310        320        330        340        350
HAERPPGPRR PERGPRGKGK ARASQVKPGD SLPRRGPGAT GIGTPAAGPG
      360        370        380        390        400
EERVGAAKAS RWRGRQNREK RFTFVLAVVI GVFVVCWFPF FFTYTLTAVG
      410        420        430        440        450
CSVPRTLFKF FFWFGYCNSS LNPVIYTIFN HDFRRAFKKI LCRGDRKRIV
```

Endothelin

```
10          20          30
CSCSSLMNKE CVTFCHLNII W
```

Endothelin receptor

```
        10          20          30          40          50
METLCLRASF WLALVGCVIS DNPERYSTNL SNHVDDFTTF RGTELSFLVT
        60          70          80          90          100
THQPTNLVLP SNGSMHNYCP QQTKITSAFK YINTVISCTI FIVGMVGNAT
        110         120         130         140         150
LLRIIYQNKC MRNGPNALIA SLALGDLIYV VIDLPINVFK LLAGRWPFDH
        160         170         180         190         200
NDFGVFLCKL FPFLQKSSVG ITVLNLCALS VDRYRAVASW SRVQGIGIPL
        210         220         230         240         250
VTAIEIVSIW ILSFILAIPE AIGFVMVPFE YRGEQHKTCM LNATSKFMEF
        260         270         280         290         300
YQDVKDWWLF GFYFCMPLVC TAIFYTLMTC EMLNRRNGSL RIALSEHLKQ
        310         320         330         340         350
RREVAKTVFC LVVIFALCWF PLHLSRILKK TVYNEMDKNR CELLSFLLLM
        360         370         380         390         400
DYIGINLATM NSCINPIALY FVSKKFKNCF QSCLCCCCYQ SKSLMTSVPM
        410         420
NGTSIQWKNH DQNNHNTDRS SHKDSMN
```

Appendix III:MATLAB code for characterization

```
close all; clear all;clc;
%% 1. for a.data
a = importdata('800 ng-mL frame 30 dish 7.xlsx');
a.data = a.data(~ismember(1:size(a.data, 1), [1:5]), :);         %rows
a.data( :, ~any(a.data,1) ) = [];                          %columns
runningfile_a = a.data(1:end,:);
time0 = a.data(:,2);
time_a = time0/1000;
all_cell_Ca_response = runningfile_a(:,3:end);
[row_all, column_all] = size(all_cell_Ca_response);
% %% 2. for b.data
% b = importdata('10 ng frame 30 dish 8.xlsx');
%
% b.data = b.data(~ismember(1:size(b.data, 1), [1:5]), :);         %rows
% b.data( :, ~any(b.data,1) ) = [];                          %columns
%
% runningfile_b = b.data(1:end,:);
% time00 = b.data(:,2);
% time_b = time00/1000;
%
% calresponse_b = runningfile_b(:,3:end);
% [row_b,column_b] = size(calresponse_b);
%% 3. Interpolation
% g=1;
% for i=1:1:column_b
%    runningbase = calresponse_b(:,i);
%    for j=1:1:length(time_b)
%      calresponse_interp(j) = interp1(time_b,calresponse_b(:,i),time_a(j),'pchip','extrap');
%    end
%
%    all_calresponse_interp(:,i) = calresponse_interp;
%
%    figure(1)
%    plot(time_b,calresponse_b(:,i)+16*g,'r');
%    hold on
%    plot(time_a,all_calresponse_interp(:,i)+16*g,'k');
%
%    axis([0 130 17 350]);
%    xlabel('Time(sec)','Fontsize',16);
%    ylabel('Fluo-4 intensity','Fontsize',16);
```

47

```matlab
%     title('Dose=10 ng (Interpolation)','Fontsize',16);
%     g=g+1;
% end
%% 4. All cell Ca response in one variable
% all_cell_Ca_response = horzcat(calresponse_a,all_calresponse_interp);
% [row_all,column_all] = size(all_cell_Ca_response);
% 5. Normalization
for i=1:1:column_all
    runningbasemeanst = all_cell_Ca_response(:,i);
    runningbasemean = sum(runningbasemeanst(1:25))/25;
    runningcalnorm = all_cell_Ca_response(:,i)/runningbasemean;
    runningcalnormnew(:,i) = runningcalnorm;
end


% 6. Appending Ca response plot after normalization with highlighted maxima in each cell response
g=1;
for i=1:1:column_all
    figure(2)
    plot(time_a,all_cell_Ca_response(:,i)+16*g,'k')
    set(gca, 'XTick',[30 130],'XGrid','on');
    set(gca, 'YTickLabel',{1:1:column_all}, 'YTick',[16:16:448] ,'Fontsize',8);
     hold on
    m=max(all_cell_Ca_response(:,i));
    t=time_a(find(all_cell_Ca_response(:,i)==m));
    plot(t,m+16*g,'r*');
        axis([0 130 15 450]);
    xlabel('Time(sec)','Fontsize',13);
    ylabel('Fluo-4 intensity','Fontsize',13);
    title('Dose=10 ng','Fontsize',13);
        hold on
    g=g+1;
end


%% 7. Finding Ca_max and corresponding T_max in each individual cell
n=1;
for i=1:1:column_all
    [max_response(:,n), I(:,n)] = max(all_cell_Ca_response(:,i));
    tmax(:,n)=time_a(find(all_cell_Ca_response(:,i)==max_response(:,n)));
        n=n+1;
end
%% 8. Histogram
% For Ca_max
inval1 =2;
w1 =1;
len1 = length(max_response);
N1=max(max_response);
x1 = 0:inval1:N1;
%plotting the histogram
figure(3)
[countlow1,cenlow1] = hist(max_response);
% Cells in percent
bar(cenlow1,(countlow1*100)/len1);
perlow1 =(countlow1*100)/len1;
histspike1(:,1)= cenlow1;
histspike1(:,2) = perlow1;
%axis([0 30 0 40]);
xlabel('Maximum Fluo-4 intensity','Fontsize',16);
ylabel('Percentage of cells','Fontsize',16);
title('Dose=10 ng','Fontsize',16);
```

48

```matlab
% For T_max
inval2 = 12;
w2 =6;
len2 = length(tmax);
N2=max(tmax);
x2 = 0:inval2:N2;
%plotting the histogram
figure(4)
[countlow2,cenlow2] = hist(tmax,x2);
% Cells in percent
bar(cenlow2,(countlow2*100)/len2);
perlow2 =(countlow2*100)/len2;
histspike2(:,1)= cenlow2;
histspike2(:,2) = perlow2;
%axis([0 120 0 45]);
xlabel('Corresponding maxima time(sec)','Fontsize',16);
ylabel('Percentage of cells','Fontsize',16);
title('Dose=10 ng','Fontsize',16);
%% 9. Clustering based on Ca_max and T_max by k-means
value = [max_response', tmax'];
k=3;
[idx1,C1]= kmeans(value,k,'Replicates',100);
% hmo = HeatMap(idx1);
location1 = knnsearch(value,C1);
figure(5)
plot(value(idx1==1,1),value(idx1==1,2),'ro')
hold on
plot(value(idx1==2,1),value(idx1==2,2),'bo')
hold on
plot(value(idx1==3,1),value(idx1==3,2),'co')
hold on
plot(value(idx1==4,1),value(idx1==4,2),'mo')
plot(C1(:,1),C1(:,2),'kx','MarkerSize',12,'LineWidth',3)
%axis([0 30 0 140]);
xlabel('Maximum Fluo-4 intensity','Fontsize',16);
ylabel('Corresponding maxima time(sec)','Fontsize',16);
title('Dose=10 ng','Fontsize',16);
figure(19);
silhouette(value,idx1);
%% 14. Tracking back the cells on the basis of PCA and k-means clustering
g=1;
for i=1:1:column_all
    figure(6)
        if idx1(i)==1
        plot(time_a,all_cell_Ca_response(:,i)+16*g,'r');
    elseif idx1(i)==2
        plot(time_a,all_cell_Ca_response(:,i)+16*g,'b');
    elseif idx1(i)==3
        plot(time_a,all_cell_Ca_response(:,i)+16*g,'c');
    else
        plot(time_a,all_cell_Ca_response(:,i)+16*g,'m');
    end
    set(gca, 'XTick',[30 130],'XGrid','on');
    set(gca, 'YTickLabel',{1:1:column_all}, 'YTick',[16:16:448] ,'Fontsize',8);
        axis([0 130 15 450]);
    xlabel('Time(sec)','Fontsize',13);
    ylabel('Fluo-4 intensity','Fontsize',13);
    title('Dose=10 ng','Fontsize',13);
        hold on
```

```matlab
        g=g+1;
    %annotation('textarrow',[0.309 0.309],[0.96 0.925],'String','SDF-1\alpha')
end


%% Plots for the representative cluster (clustering based on Ca_max vs T_max)
figure(45)
subplot(3,1,1)
plot (time_a,runningcalnormnew(:,location1(3,1)),'c');
xlabel('Time(sec)','Fontsize',13);
ylabel('Fluo-4 intensity','Fontsize',13);
annotation('textbox',[0.68 0.87 0.32 0.06],'String',{'\bf S1'},'FitBoxToText','off','LineStyle','none');
%title('S1','Fontsize',13);
axis([0 130 0 3.2]);
subplot(3,1,2)
plot (time_a,runningcalnormnew(:,location1(2,1)),'b');
xlabel('Time(sec)','Fontsize',13);
ylabel('Fluo-4 intensity','Fontsize',13);
annotation('textbox',[0.68 0.57 0.32 0.06],'String',{'\bf S2'},'FitBoxToText','off','LineStyle','none');
%title('S2','Fontsize',13);
axis([0 130 0 3.2]);
subplot(3,1,3)
plot (time_a,runningcalnormnew(:,location1(1,1)),'r');
xlabel('Time(sec)','Fontsize',13);
ylabel('Fluo-4 intensity','Fontsize',13);
annotation('textbox',[0.68 0.27 0.32 0.06],'String',{'\bf S3'},'FitBoxToText','off','LineStyle','none');
%title('S3','Fontsize',13);
axis([0 130 0 3.2]);
%% 10. Stacked bar plot by Ca_max and T_max and k-means
xpCaT = length(idx1);
% For IDX=1
    j=1;
    for i = 1:1:xpCaT
    if idx1(i)==1
        SPCaTa(j)=idx1(i);      % the cells in cluster 1
        j=j+1;
    else
    end
    end
    % For IDX=2
    j=1;
    for i = 1:1:xpCaT
    if idx1(i)==2
        SPCaTb(j)=idx1(i);      % the cells in cluster 2
        j=j+1;
    else
    end
    end

    % For IDX=3
    j=1;
    for i = 1:1:xpCaT
    if idx1(i)==3
        SPCaTc(j)=idx1(i);      % the cells in cluster 3
        j=j+1;
    else
    end
    end
    % Subpopulation percentage calculation
SPCaT1=(length(SPCaTa)/xpCaT)*100;
```

```
SPCaT2=(length(SPCaTb)/xpCaT)*100;
SPCaT3=(length(SPCaTc)/xpCaT)*100;
pyCaT=[SPCaT1, SPCaT2,SPCaT3];
figure(7)
hbarCaT=bar([pyCaT; nan(1,3)], 'Stacked');
set(gca, 'YTickLabel',num2str(1.*get(gca,'YTick')','%g%%'),'Fontsize',14)
axis([0.8 1 0 100])
set(gca,'xtick',1)
set(hbarCaT,{'FaceColor'},{'r';'b';'c'});
xt = get(gca, 'XTick');
set(gca, 'XTick', xt, 'XTickLabel', {''});
yd = get(hbarCaT, 'YData');
labels = {'S3','S2','S1'};
barbase = cumsum([zeros(size(pyCaT,1),1) pyCaT(:,1:end-1)],2);
joblblpos = pyCaT/2 + barbase;
for k1 = 1:size(pyCaT,1)
    text(xt(k1)*ones(1,size(pyCaT,2)), joblblpos(k1,:), labels, 'HorizontalAlignment','Right','Fontsize',12);
end
title('Dose=10 ng','Fontsize',16);
%legend(hbarCaT, {'SP1=64.29%', 'SP2=35.71%', 'SP2=35.71%', 'SP2=35.71%'},'Location','southoutside');
legend('boxoff')
%% 11. Area under curve calculation
for i=1:1:column_all
    IT=I';
    a=IT(i);
    y=all_cell_Ca_response(:,i);
    yi=[];
    timei=[];
    for xi=1:1:a
        yi(xi)=y(xi);
        timei(xi)=time_a(xi);
    end
    yi;
    timei;
        area=trapz(timei,yi);
        tot_a(i)=area;
end
%% 12. Histogram for area under curve
inval3 = 60;
w3 = 30;
len3 = length(tot_a);
N3=max(tot_a);
x3 = 0:inval3:N3;
%plotting the histogram
figure(8)
[countlow3,cenlow3] = hist(tot_a,x3);
% Making the Cells in percent
bar(cenlow3,(countlow3*100)/len3);
perlow3 =(countlow3*100)/len3;
histspike3(:,1)= cenlow3;
histspike3(:,2) = perlow3;
%axis([-1 1300 0 25]);
xlabel('Area under curve(arbitrary units)','Fontsize',16);
ylabel('Percentage of cells','Fontsize',16);
title('Dose=10 ng','Fontsize',16);
%% 13. PCA and k-means by New code
cov_lem = cov(all_cell_Ca_response);
lem = eig(cov_lem);
lem_s = sort(lem,'descend');
```

```
% figure(9)
% plot(lem_s);
% xlabel('Number of eigen values','Fontsize',16);
% ylabel('Eigen value(\lambda)','Fontsize',16);
% title('Dose=10 ng','Fontsize',16);
%set(gca, 'XTick',[0:1:rown_all])
% for saving the values for ksdensity plot
% all_in_one=vertcat(max_response, tmax, tot_a);
% for_ksdensity_eig=horzcat(lem_s,all_in_one');
% PCA
noeig=8;
[pcomp, transdata]= pcanalysis(all_cell_Ca_response', noeig);
% k-means
% k=2;
[idx2,C2]= kmeans(transdata,k,'Replicates',100);
% to visualize in 2D
pc=2;
[pcomp1, transdata1]= pcanalysis(transdata, pc);
% New centroid position
New_centroid = C2*pcomp1;
location2 = knnsearch(transdata1,New_centroid);
cluster1 = transdata1(idx2 == 1, :);
cluster2 = transdata1(idx2 == 2, :);
cluster3 = transdata1(idx2 == 3, :);
cluster4 = transdata1(idx2 == 4, :);
figure(10);
scatter(cluster1(:,1),cluster1(:,2),'ro');
hold on;
scatter(cluster2(:,1),cluster2(:,2),'bo');
hold on;
scatter(cluster3(:,1),cluster3(:,2),'co');
hold on;
scatter(cluster4(:,1),cluster4(:,2),'mo');
hold on;
plot(New_centroid(:,1),New_centroid(:,2),'kx','MarkerSize',12,'LineWidth',3);
%axis([-40 40 0 180]);
xlabel('PC1','Fontsize',16);
ylabel('PC2','Fontsize',16);
title('Dose=10 ng','Fontsize',16);
%% 14. Tracking back the cells on the basis of PCA and k-means clustering
g=1;
for i=1:1:column_all
    figure(11)
        if idx2(i)==1
        plot(time_a,all_cell_Ca_response(:,i)+16*g,'r');
    elseif idx2(i)==2
        plot(time_a,all_cell_Ca_response(:,i)+16*g,'b');
    elseif idx2(i)==3
        plot(time_a,all_cell_Ca_response(:,i)+16*g,'c');
    else
        plot(time_a,all_cell_Ca_response(:,i)+16*g,'m');
    end
    set(gca, 'XTick',[30 130],'XGrid','on');
    set(gca, 'YTickLabel',{1:1:column_all}, 'YTick',[16:16:448] ,'Fontsize',8);
    axis([0 130 15 450]);
    xlabel('Time(sec)','Fontsize',13);
    ylabel('Fluo-4 intensity','Fontsize',13);
    title('Dose=10 ng','Fontsize',13);
    hold on
```

```
        g=g+1;
end
%% Plots for the representative cluster (clustering based on K-means)
figure(55)
subplot(3,1,1)
plot (time_a,runningcalnormnew(:,location2(3,1)),'c');
xlabel('Time(sec)','Fontsize',13);
ylabel('Fluo-4 intensity','Fontsize',13);
annotation('textbox',[0.68 0.87 0.32 0.06],'String',{'\bf S1'},'FitBoxToText','off','LineStyle','none');
%title('S1','Fontsize',13);
axis([0 130 0 3.2]);
subplot(3,1,2)
plot (time_a,runningcalnormnew(:,location2(2,1)),'b');
xlabel('Time(sec)','Fontsize',13);
ylabel('Fluo-4 intensity','Fontsize',13);
annotation('textbox',[0.68 0.57 0.32 0.06],'String',{'\bf S2'},'FitBoxToText','off','LineStyle','none');
%title('S2','Fontsize',13);
axis([0 130 0 3.2]);
subplot(3,1,3)
plot (time_a,runningcalnormnew(:,location2(1,1)),'r');
xlabel('Time(sec)','Fontsize',13);
ylabel('Fluo-4 intensity','Fontsize',13);
annotation('textbox',[0.68 0.27 0.32 0.06],'String',{'\bf S3'},'FitBoxToText','off','LineStyle','none');
%title('S3','Fontsize',13);
axis([0 130 0 3.2]);
%% 15. Stacked bar plot by PCA and k-means
xp = length(idx2);
% for idx2=1
  j=1;
  for i = 1:1:xp
  if idx2(i)==1
     SPa(j)=idx2(i);      % the cells in cluster 1
     j=j+1;
  else
  end
  end

% for idx2=2
  j=1;
  for i = 1:1:xp
  if idx2(i)==2
     SPb(j)=idx2(i);      % the cells in cluster 2
     j=j+1;
  else
  end
  end
  % for idx2=3
  j=1;
  for i = 1:1:xp
  if idx2(i)==3
     SPc(j)=idx2(i);      % the cells in cluster 3
     j=j+1;
  else
   end
  end
SP1=(length(SPa)/xp)*100;
SP2=(length(SPb)/xp)*100;
SP3=(length(SPc)/xp)*100;
py=[SP1, SP2, SP3];
```

```
figure(12)
hbar=bar([py; nan(1,3)], 'Stacked');
set(gca, 'YTickLabel',num2str(1.*get(gca,'YTick')','%g%%'),'Fontsize',14);
axis([0.8 1 0 100]);
set(hbar,{'FaceColor'},{'r';'b';'c'});
xt = get(gca, 'XTick');
set(gca, 'XTick', xt, 'XTickLabel', {''});
yd = get(hbar, 'YData');
labels = {'S3','S2','S1'};
barbase = cumsum([zeros(size(py,1),1) py(:,1:end-1)],2);
joblblpos = py/2 + barbase;
for k1 = 1:size(py,1)
    text(xt(k1)*ones(1,size(py,2)), joblblpos(k1,:), labels, 'HorizontalAlignment','Right','Fontsize',12);
end
title('Dose=10 ng','Fontsize',16);
%legend(hbar, {'SP1=87.50%','SP2=12.50%','SP3=12.50%','SP4=12.50%'},'Location','southoutside');
legend('boxoff')
%% 16. Silhoutte plot
figure(13);
silhouette(transdata,idx2);
title('Dose=10 ng','Fontsize',16);
```