

# STRUCTURAL VARIATION DISCOVERY IN REPEAT- AND DUPLICATION-RICH REGIONS

A THESIS SUBMITTED TO  
THE GRADUATE SCHOOL OF ENGINEERING AND SCIENCE  
OF BILKENT UNIVERSITY  
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR  
THE DEGREE OF  
MASTER OF SCIENCE  
IN  
COMPUTER ENGINEERING

By  
Emre Dogru  
February 2018

STRUCTURAL VARIATION DISCOVERY IN REPEAT- AND  
DUPLICATION-RICH REGIONS

By Emre Dogru

February 2018

We certify that we have read this thesis and that in our opinion it is fully adequate,  
in scope and in quality, as a thesis for the degree of Master of Science.

---

Can Alkan(Advisor)

---

No academic titles, only the name Y

---

No academic titles, only the name Z

Approved for the Graduate School of Engineering and Science:

---

Ezhan Karaşan  
Director of the Graduate School

# ABSTRACT

## STRUCTURAL VARIATION DISCOVERY IN REPEAT- AND DUPLICATION-RICH REGIONS

Emre Dogru

M.S. in Computer Engineering

Advisor: Can Alkan

February 2018

Write the abstract here.

*Keywords:* Key one, key two.

## ÖZET

# TEKRARLI VE DUPLİKASYON BAKIMINDAN ZENGİN BÖLGELERDE YAPISAL VARYASYON KEŞFİ

Emre Dogru

Bilgisayar Mühendisliği, Yüksek Lisans

Tez Danışmanı: Can Alkan

Şubat 2018

Türkçe özet buraya gelecek.

*Anahtar sözcükler:* Anahtar Sözcüklerim.

# Acknowledgement

I acknowledge that ...

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Sequencing technologies . . . . .	2
1.1.1	Early Sequencing Methods . . . . .	2
1.1.2	High-Throughput Sequencing . . . . .	2
1.2	Genomic Variation . . . . .	2
1.2.1	Structural Variation . . . . .	2
1.3	Segmental Duplication . . . . .	2
1.3.1	Singly Unique Nucleotides . . . . .	2
1.4	Our Contribution . . . . .	2
<b>2</b>	<b>Gene Copy Number</b>	<b>3</b>
2.1	Data Preprocessing . . . . .	4
2.2	Extracting Singly Unique Nucleotides . . . . .	4
2.3	Read Depth . . . . .	4

2.3.1	GC Correction . . . . .	4
2.4	Results . . . . .	4
3	Paralog Gene Copy Number	5
3.1	Motivation . . . . .	5
3.2	Results . . . . .	5
4	Conclusion and Discussion	6
4.1	Future Work . . . . .	6
A	Data	8
B	Code	9

# List of Figures



# List of Tables

# Chapter 1

## Introduction

Human genome project ...

## 1.1 Sequencing technologies

### 1.1.1 Early Sequencing Methods

### 1.1.2 High-Throughput Sequencing

## 1.2 Genomic Variation

### 1.2.1 Structural Variation

#### 1.2.1.1 Copy Number Variation

## 1.3 Segmental Duplication

### 1.3.1 Singly Unique Nucleotides

## 1.4 Our Contribution

## Chapter 2

### Gene Copy Number

## **2.1 Data Preprocessing**

## **2.2 Extracting Singly Unique Nucleotides**

## **2.3 Read Depth**

### **2.3.1 GC Correction**

## **2.4 Results**

## Chapter 3

# Paralog Gene Copy Number

### 3.1 Motivation

### 3.2 Results

## Chapter 4

# Conclusion and Discussion

### 4.1 Future Work

# Bibliography



# Appendix A

## Data

# Appendix B

## Code