Pipelined Methylation Analysis and Mapping Reads for Display on IGV
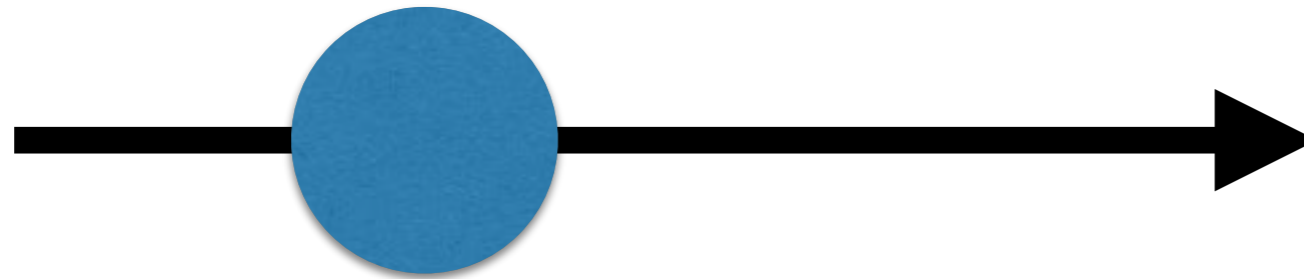
Pat Johnson, Alister Maguire, Hiranmayi Duvvuri
Vince Bicocca
University of Oregon
CIS 454: Bioinformatics

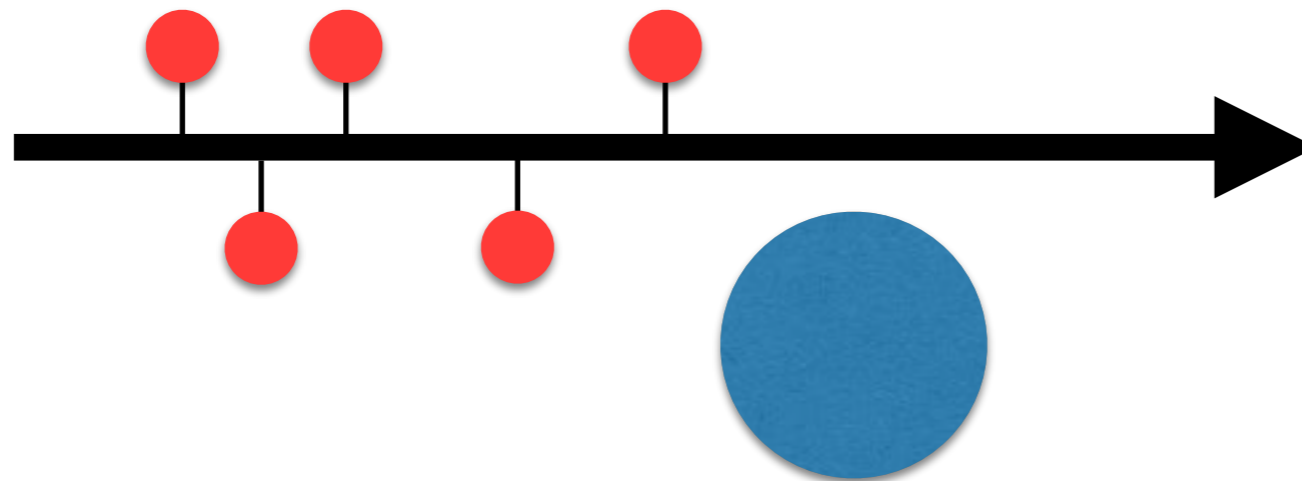# Cytosine methylation can alter temporal and spatial gene expression

No methylation

Active gene expression

Transcription factors can bind

Methylation
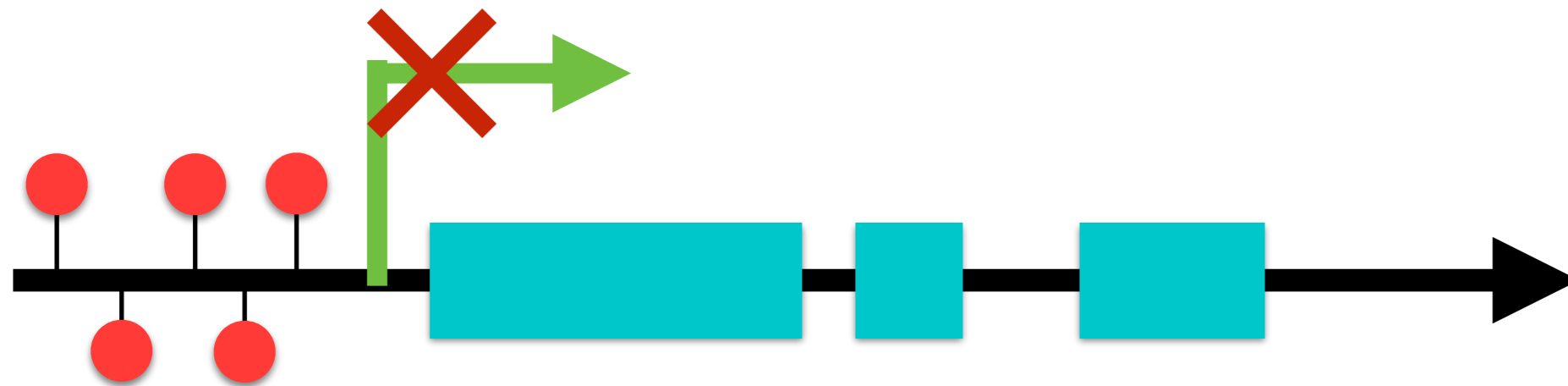

Repressed gene expression

Transcription factors can not bind

Defects in meythlation can give rise to detrimental diseases
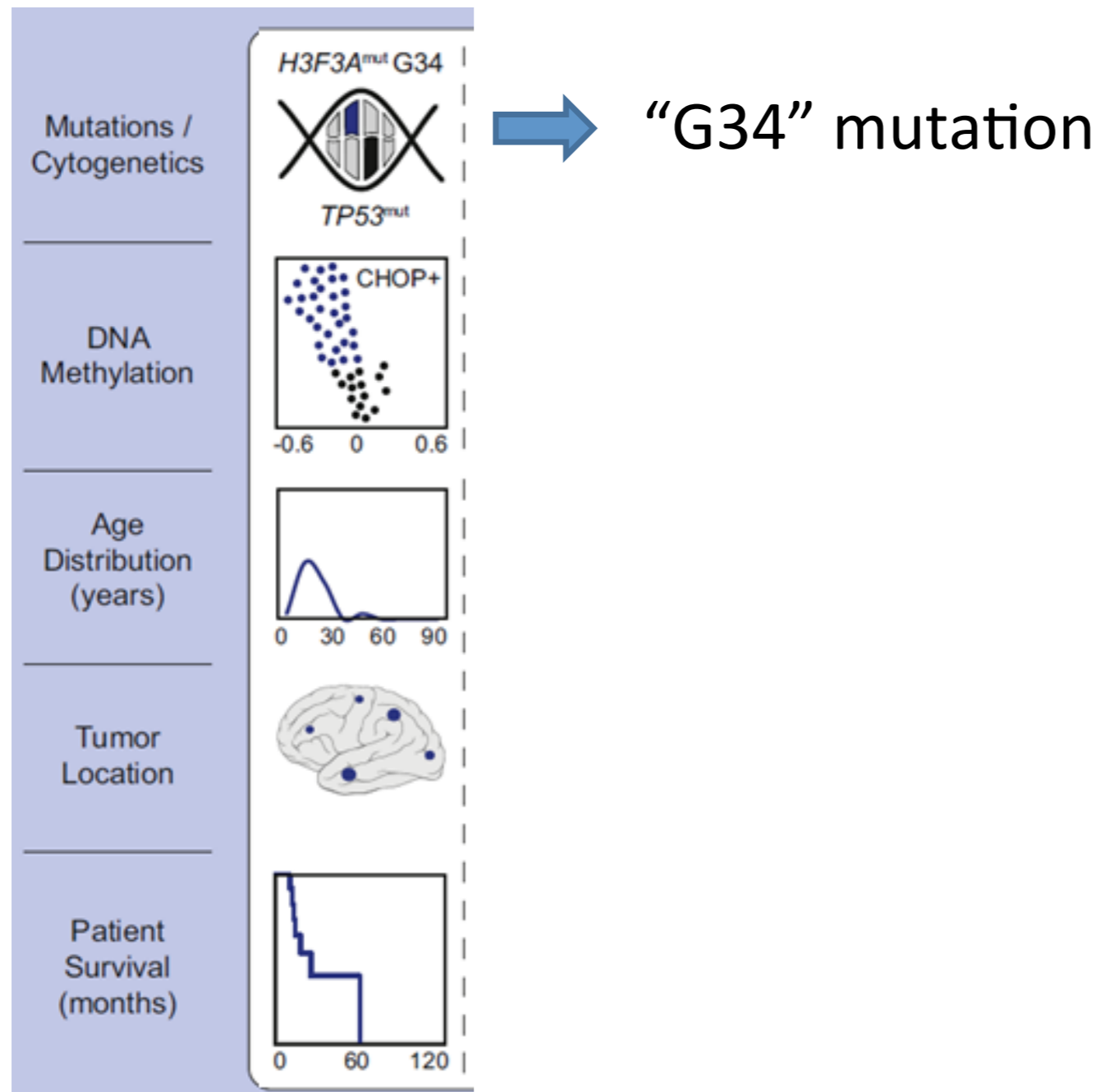
Tumor Suppressor Gene

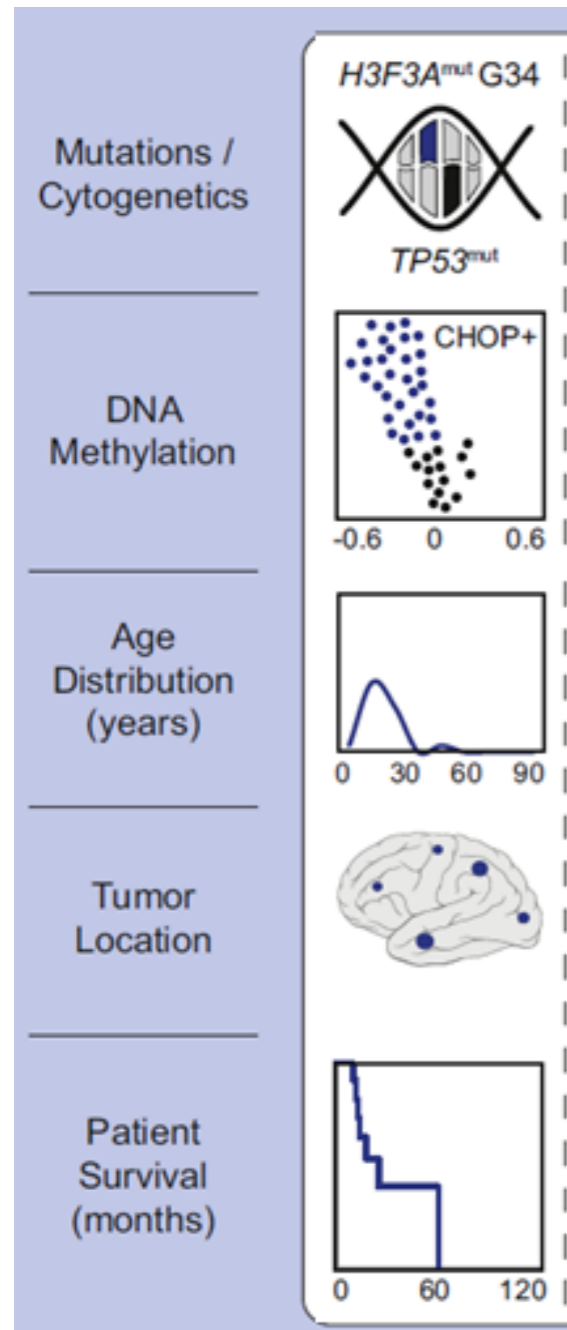Controlled Proliferation

Tumorigenesis

A type of childhood brain tumor has unique characteristic as a result of a shared mutation

→ "G34" mutation

A type of childhood brain tumor has unique characteristic as a result of a shared mutation

"G34" mutation

Abnormally low DNA methylation (hypomethylation)

Sturm et al. *Cancer Cell*. 2012

A type of childhood brain tumor has unique characteristic as a result of a shared mutation

Mutations / Cytogenetics — H3F3A^mut G34, TP53^mut → "G34" mutation

DNA Methylation — CHOP+ (-0.6, 0, 0.6) → Abnormally low DNA methylation (hypomethylation)

Age Distribution (years) (0, 30, 60, 90) → Peak onset around 15 years old

Tumor Location

Patient Survival (months) (0, 60, 120)
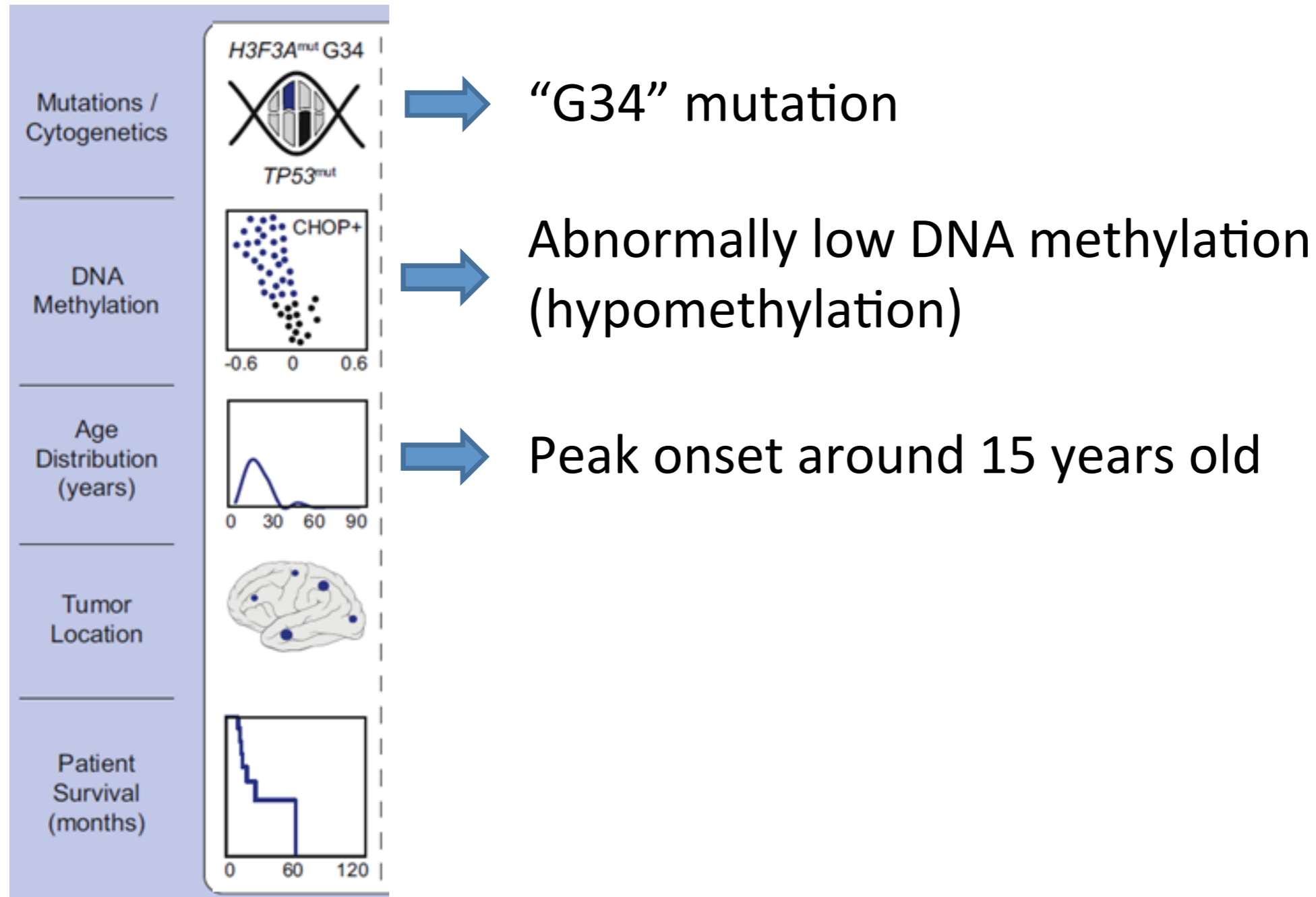
Sturm et al. *Cancer Cell*. 2012

A type of childhood brain tumor has unique characteristic as a result of a shared mutation

"G34" mutation

Abnormally low DNA methylation (hypomethylation)

Peak onset around 15 years old

Tumors arise in outer lobes of the brain

A type of childhood brain tumor has unique characteristic as a result of a shared mutation

Mutations / Cytogenetics
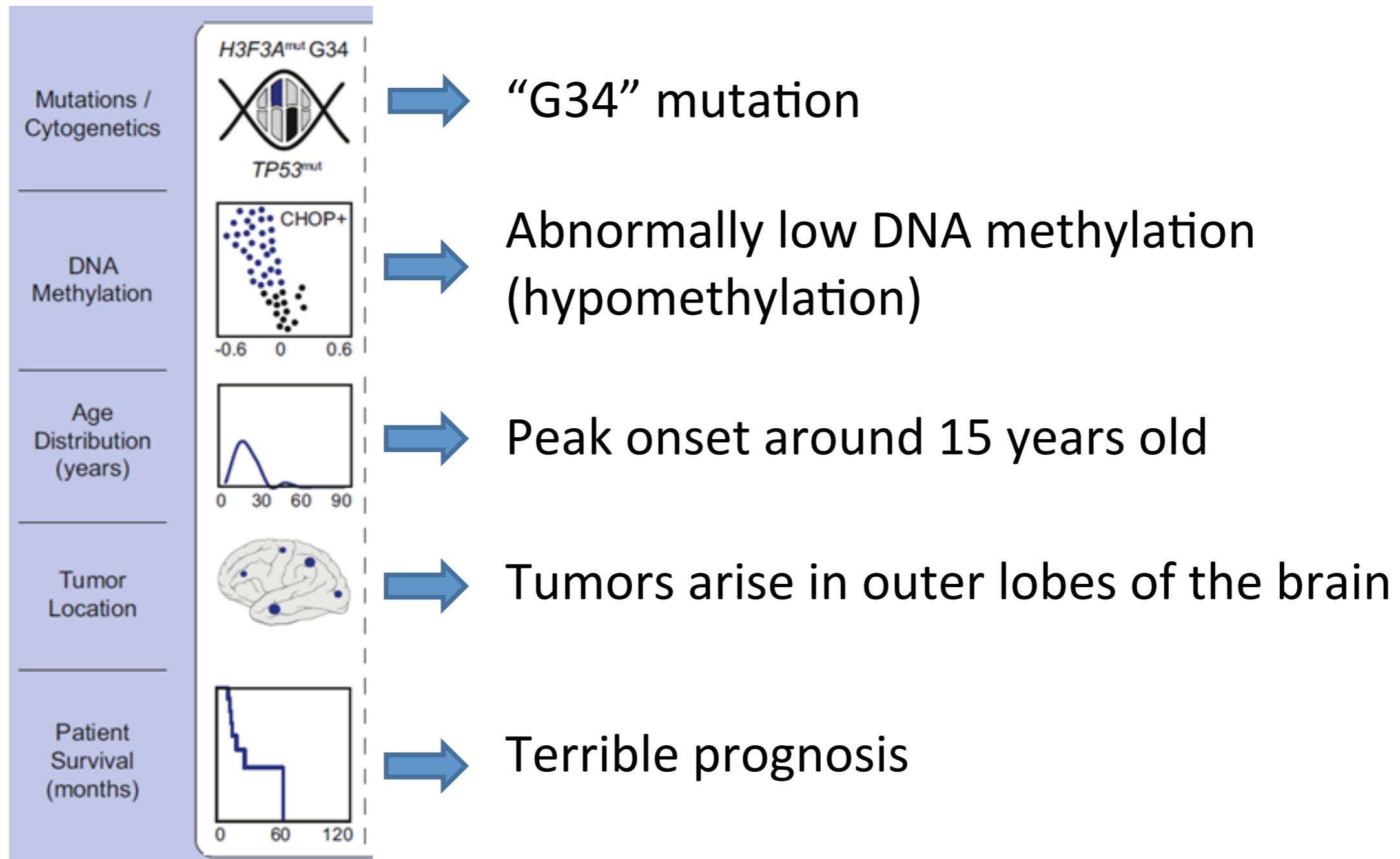H3F3A^mut G34
TP53^mut
→ "G34" mutation

DNA Methylation
CHOP+
-0.6   0   0.6
→ Abnormally low DNA methylation (hypomethylation)

Age Distribution (years)
0   30   60   90
→ Peak onset around 15 years old

Tumor Location
→ Tumors arise in outer lobes of the brain

Patient Survival (months)
0   60   120
→ Terrible prognosis

Sturm et al. *Cancer Cell*. 2012

A type of childhood brain tumor has unique characteristic as a result of a shared mutation

Mutations / Cytogenetics	*H3F3A*ᵐᵘᵗ G34 *TP53*ᵐᵘᵗ	→ "G34" mutation
DNA Methylation	CHOP+ -0.6 0 0.6	→ Abnormally low DNA methylation (hypomethylation)
Age Distribution (years)	0 30 60 90	→ Peak onset around 15 years old
Tumor Location		→ Tumors arise in outer lobes of the brain
Patient Survival (months)	0 60 120	→ Terrible prognosis

- How this mutation results in cancer is not understood.
- A better understanding of how the tumors develop should result in better treatments.

Sturm et al. *Cancer Cell*. 2012

Illumina Sequencing based readouts of epigenetic modifications can measure DNA accessibility and modification

1) Whole genome sequencing
   - Identify genome wide DNA methylation

2) Chromatin Immunoprecipitation-Seq (ChiP-Seq)
   - Identify genome-wide DNA binding sites for transcription factors and other proteins
   - Identify histone modifications
   - Reveal gene regulator networks in combination with RNA sequencing and methylation analysis
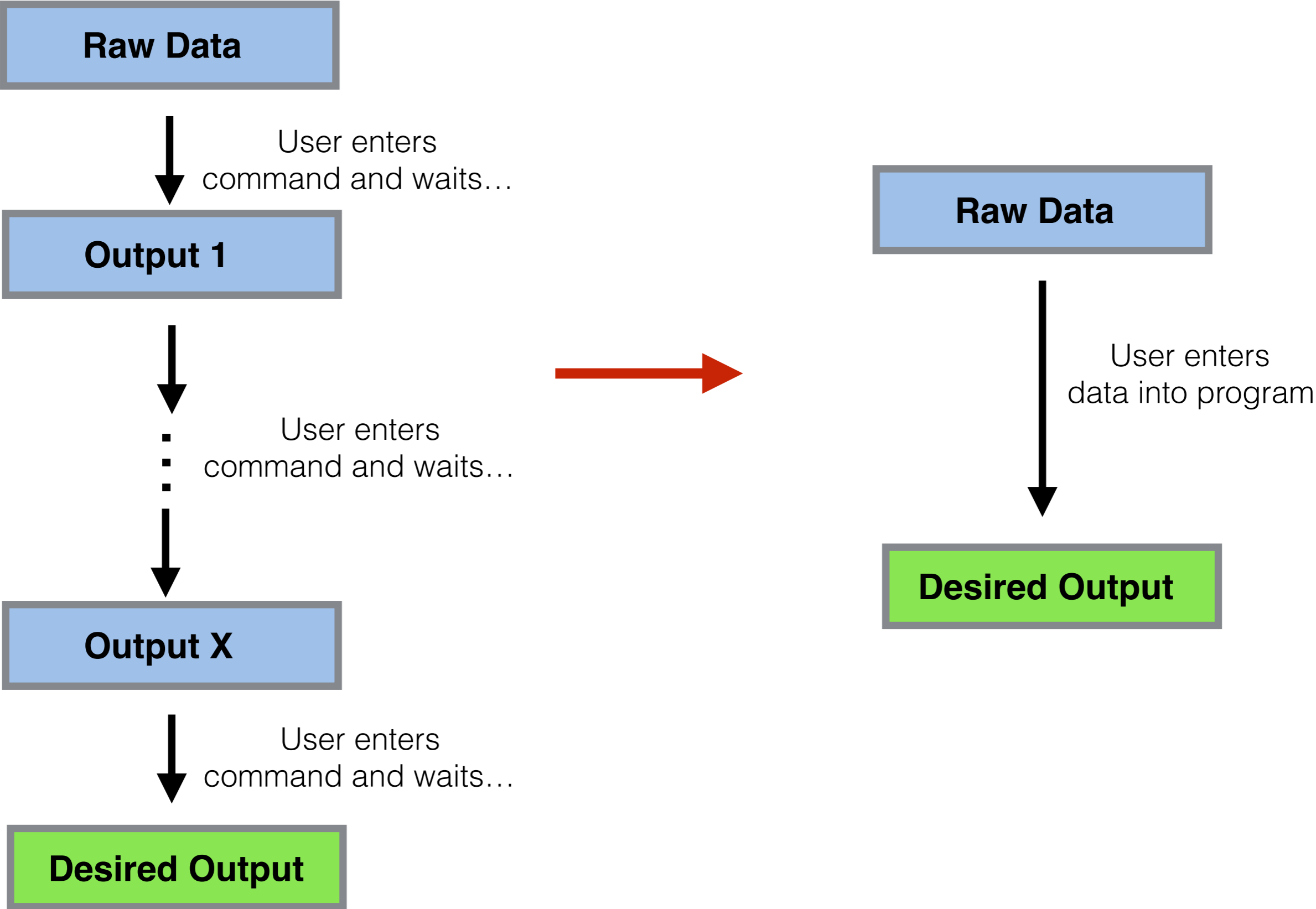
Illumina Sequencing based readouts of epigenetic modifications can measure DNA accessibility and modification

1) Whole genome sequencing
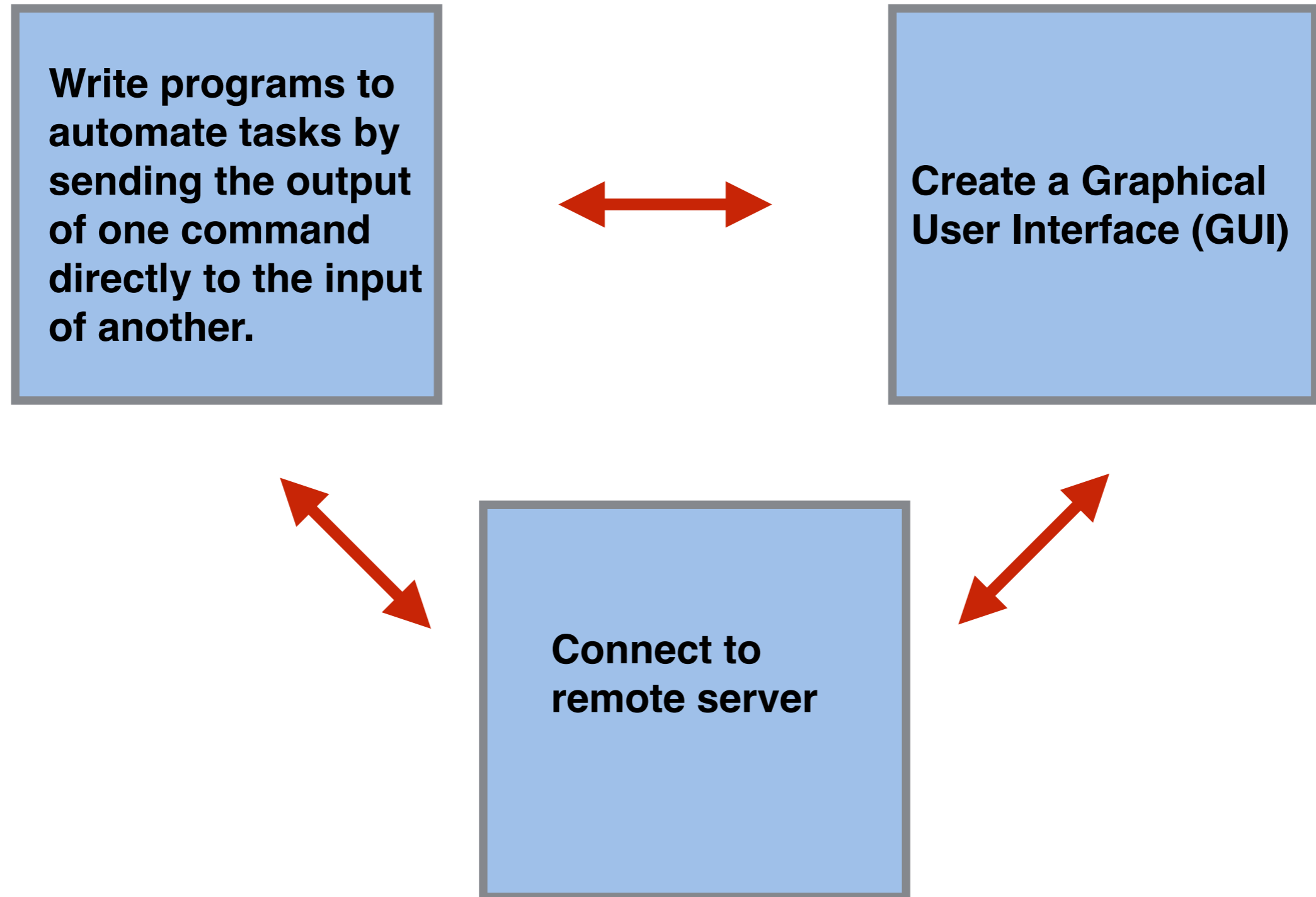   - Identify genome wide DNA methylation

2) Chromatin Immunoprecipitation-Seq (ChiP-Seq)
   - Identify genome-wide DNA binding sites for transcription factors and other proteins
   - Identify histone modifications
   - Reveal gene regulator networks in combination with RNA sequencing and methylation analysis

*Lots of Data…time consuming and tedious to convert into usable results

Goal: Create a user-friendly interface to automate and speed up the process

Raw Data

↓ User enters command and waits…

Output 1

⋮ User enters command and waits…

Output X

↓ User enters command and waits…

Desired Output

→

Raw Data

↓ User enters data into program

Desired Output

Project Outline

Write programs to automate tasks by sending the output of one command directly to the input of another.

Create a Graphical User Interface (GUI)

Connect to remote server

Challenges Along the Way

unfamiliar protocol => look up documentation

making the protocols as user friendly as possible

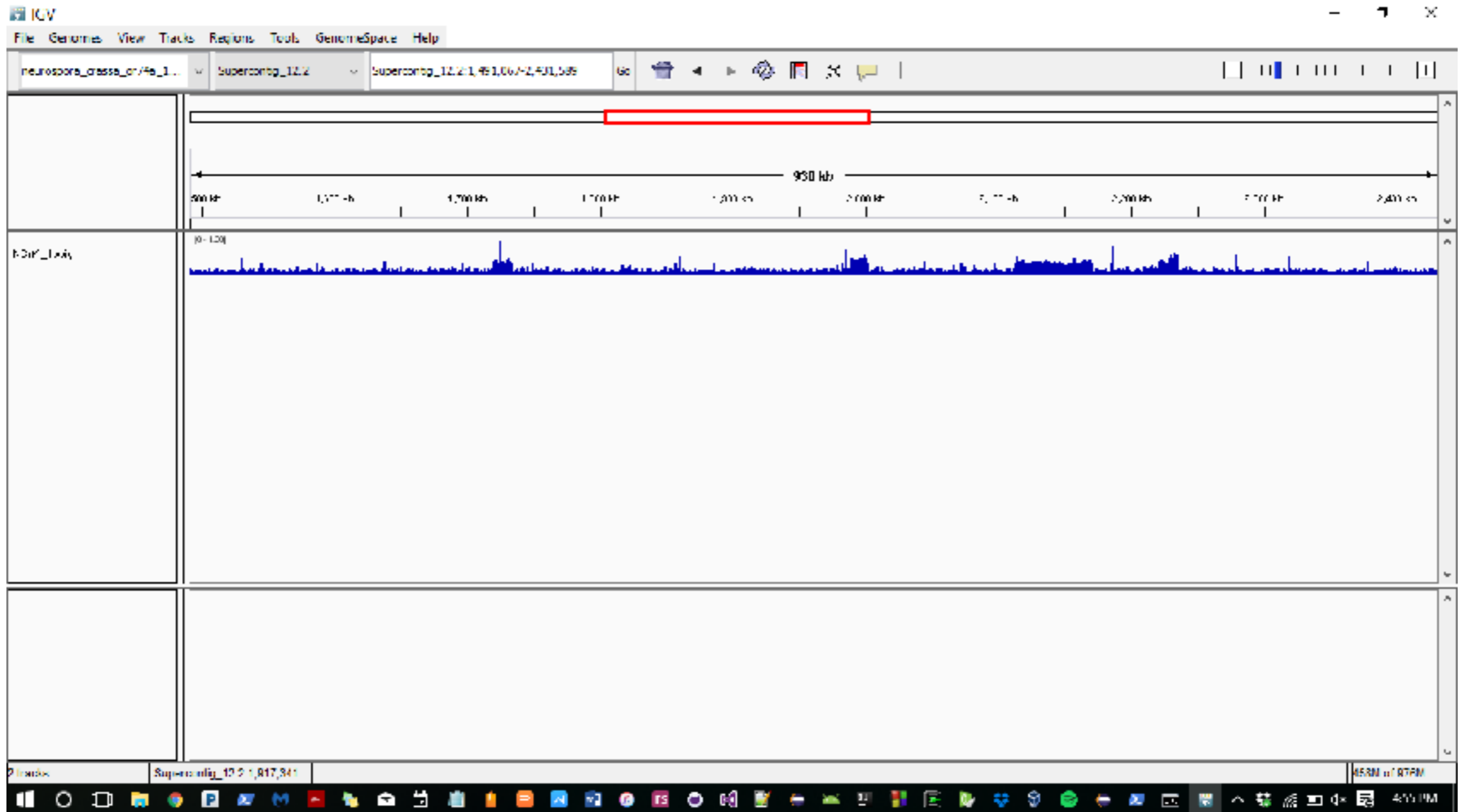merging our pipelines into a single GUI while retaining independence

need for new modules (ttk, paramiko)

unexpected ACISS module behaviors

protocol commands lacking documentation

missing files/directories

Demo

If you get stuck using the programs…

1) There will be documentation loaded onto your computer

2) Our contact information (email) is in the documentation