**Aneuvis: Reproducible web-based visualization of aneuploidy in single cells**

**Aneuvis: Web-based exploration of numerical chromosomal variation in single cells**

Daniel G. Piqué1,2, Samuel E. Zimmerman1, Grasiella A. Andriani2, Elaine Maggi2, John M. Greally2, Cristina Montagna2,3, and Jessica C. Mar1,4,5\*

1Department of Systems and Computational Biology

2Department of Genetics

3Department of Pathology

4Department of Epidemiology and Population Health

Albert Einstein College of Medicine, Bronx, NY, 10461, USA

5Australian Institute for Bioengineering and Nanotechnology, The University of Queensland, QLD 4072, Australia

***Bioinformatics* Application Notes (up to 2 pages; this is approx. 1,300 words or 1,000 words plus one figure):**Applications Notes are short descriptions of novel software or new algorithm implementations, databases and network services (web servers, and interfaces). Software or data must be freely available to non-commercial users. Availability and Implementation must be clearly stated in the article. Authors must also ensure that the software is available for a full TWO YEARS following publication. Web services must not require mandatory registration by the user. Additional Supplementary data can be published online-only by the journal. This supplementary material should be referred to in the abstract of the Application Note. If describing software, the software should run under nearly all conditions on a wide range of machines. Web servers should not be browser specific. Application Notes must not describe trivial utilities, nor involve significant investment of time for the user to install. The name of the application should be included in the title.

**Abstract**

Aneuploidy is a predisposing factor for cancer. There are traditionally two main methods for detecting aneuploidy in research and clinical practice. Molecular cytogenetic approaches, often used in clinical practice, involve direct visualization of interphase or metaphase preparations of chromosomes using fluorescent probes. More recently, whole genome sequencing has been utilized to discern chromosomal copy number at high resolution. Each of these approaches has their advantages and disadvantages, and an integrated analysis would aid in the better understanding of the strengths and limitations of these methods. Here, we develop a web-based tool, termed Aneuvis, that allows users to upload molecular cytogenetic or processed whole-genome sequencing data in a cell-by-chromosome format. Aneuvis automatically constructs visualizations and summary statistics to determine whether significant copy number variation exists between experimental treatment groups. Aneuvis is a user-friendly tool that will help researchers identify stressors and molecular perturbations that drive aneuploidy.

will help highlight chromosomal copy number heterogeneity both within and between different groups and

distinguish cell populations.

**Introduction**

In humans, alterations in chromosome number are a source of congenital disease and a predisposing factor for carcinogenesis. Increased variability in the karyotypes of cancer cell populations is also associated with resistance to chemotherapy and poor clinical outcomes. Alterations in chromosome number is also a sign of chromosomal instability, which can lead to cancer therapeutic resistance. The precise mechanisms of chromosomal instability and aneuploidy are under investigation and could lead to a better understanding of oncogenesis and cellular aging.

Single cell methods have

The various approaches to measuring and quantifying the numerical chromosomal state of a cell each have their advantages and disadvantages. Single cell whole genome DNA sequencing is touted as the next generation approach to quantifying aneuploidy at a genome-wide scale, yet it suffers from sensitivity issues. Fluorescent based approaches may be more sensitive at detecting individual chromosomal changes and, unlike with whole genome sequencing, the ploidy can be determined through direct visualization. However, microscopy-based approaches are relatively low throughput and are limited in the number of chromosomes that can be detected at one time.

One common mode of investigation into the mechanisms of aneuploidy involves a control-treatment study design, in which cells or animals are exposed to a variety of stressors, such as genetic perturbations or small molecule inhibition, followed by the documentation of the numerical variability in chromosomes within a set of single cells exposed to a given stressor. Stressors that induce more numerical chromosome aberrations can be mechanistically associated with driving aneuploidy. It is often of interest to quantify the effect of a treatment on the chromosomal state in a population of single cells, along with a measure of statistical significance for difference from a control or from other treatment states. However, there is a lack of definitive, unbiased, robust, and easy-to-use methods for biologists to draw these conclusions.

Alterations can take place as departures from an integer-valued set, known as polyploidy. There are two major methods for detecting alterations in chromosome number – fluorescent-based methods and genome sequencing.

In a cell, there are three possible states that a set of chromosomes can have. Diploidy refers to the presence of two copies of each chromosome. Polyploidy refers to an integer-values increase in the number of chromosomes, often resulting from whole-genome duplication. The presence of WGD alters sensitivity of cells to potential cancer therapeutics and is thus important to detect1,2.

Here, we introduce aneuvis, a user-friendly web application for visualizing and summarizing numerical chromosomal changes. A tutorial video is available to guide users through the interface. The application describes three different types of input that can be accepted. First, bed files that include copy number output from Ginkgo3, a web-based program for calling ploidy changes from whole-genome sequencing data, are accepted. In addition, we take multicolor fluorescent in situ hybridization data summarized at the chromosomal level. An approach using interphase FISH to estimate chromosomal copy number in various tissues was recently described by one of us4. This data is often recorded into excel spreadsheets, and the file structure accepted by aneuvis are described on the website.

Existing methodologies focus on the determination of chromosomal state from raw sequencing data and often output a cell-by-chromosome matrix containing the inferred number of chromosomes at each bin (**Figure 1b**).

Aneuvis describes takes as input .bed files, and in particular, output from Ginkgo, corresponding to copy number changes The interface guides analyzing

Instavility

Invisibility

SUMCIN CIN-VIS



**Figure 1. Overview of aneuvis workflow for analyzing variation in chromosome number.** A) The analysis of multiple cells per treatment group helps determine whether a treatment can induce changes in the number of chromosomes. B) Next, various methodologies can be used to determine the copy number state of each cell within each treatment group. This information is often stored as a cell x chromosome matrix, where the entries indicate the number of inferred copies of a chromosome in a cell. C) Aneuvis incorporates information from the experimental design as well as from chromosomal copy number matrices to determine whether differences exist between treatment groups. A table of descriptive statistics summarized by group and by chromosome is automatically generated and available for download. Visual representations of the relationship in aneuploidy between different groups is also automatically generated. A permutation-based approach enables the user to conclude whether there is a statistically significant difference in the ploidy characteristics between treatment groups.

Aneuvis sits downstream of many efforts to accurately estimate chromosomal copy number from a variety of types of data (**Figure 1**). Indeed, its role is to help categorize between-group

*Whole chromosome instability*

Whole chromosome instability is thought to arise from the missegregation of chromosomes during mitosis.

*Drivers of aneuploidy*

*Measuring aneuploidy*

Whole genome sequencing

FISH

Fluorescence in situ hybridization is an approach that uses fluorescently-labeled DNA-based probes to

Chromosomal copy number variation variation Aneuploidy is The drivers of aneuploidy are unclear. Understanding how molecular stressors are

In a cell, there are three possible states

**Results**

Aneuvis overview.

The idea behind this web application is to take in copy number data from different treatment groups and to output a summary of the relationship between different treatment groups. In addition, the same experimental treatments can be compared across multiple platforms to assess copy number (e.g. both FISH and whole genome sequencing).

There are three types of single-cell chromosomal data that can be uploaded into aneuvis. First, fluorescence in situ hybridization (FISH) data, where the chromosome number is inferred from the number of distinct fluorescent probes, can be represented as a matrix where each column is a chromosome and each row is a separate biological cell. The numbers in the matrix represent the number of copies of a given chromosome for a given cell. The second type of data is single cell whole genome sequencing data. There are currently user-friendly programs, such as Ginkgo, for converting .bam files containing aligned reads obtained from DNA-sequencing to copy number data. Aneuvis will convert output in .bed format to a summarized copy number state using a weighted mean, where the copy number at each bin contributes proportionally to its overall size. Finally, SKY data can be input into aneuvis from an Excel spreadsheet that contains chromosomal counts and alterations formatted according to the International System for human Cytogenomic Nomenclature (ISCN). Copy number information is automatically extracted from these karyotypes.

We acknowledge several limitations of aneuvis. First is the ability to incorporate sub-chromosomal events.

Analyze chromosomal counts from 2 to 4 chromosomes. Single cell whole genome sequencing (SC-WGS) analyze chromosome counts from single cell sequencing data." Spectral karyotyping (SKY)"), "- analyze chromosome counts and structural variation from all chromosomes.")

),

these datatypes. visualization and summary statistics three kinds of copy number data – single cell,

Aneuvis highlights previously unseen variation in chromosome counts

Comparing the degree of aneuploidy between predefined treatment groups

Summary statistics

Anevis compares the degree of aneuploidy between predefined treatment groups

displays new findings