**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

**Abstract**

Numerical chromosomal aneuploidy is a predisposing factor for cancer. There are two broad approaches for detecting numerical 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. An integrated, automated workflow would aid in the better understanding of the strengths and limitations of these methods. Here, we develop a web application, termed Aneuvis, that allows users to upload molecular cytogenetic or processed whole-genome sequencing data in a cell-by-chromosome matrix format. Aneuvis automatically constructs visualizations and summary statistics to determine whether significant copy number variation exists between experimental treatment groups. Permutation testing is also performed to provide the user with statistical confidence in the degree of aneuploidy between different treatment groups. Aneuvis is a user-friendly tool that will help researchers identify the genetic and environmental perturbations that promote aneuploidy.

|  |  |  |
| --- | --- | --- |
| **Method** | **Equation** | **Support** |
| Instability index ()1,2 |  |  |
| ANCA ()3,4 |  |  |
| Normalized ANCA ()\* |  |  |
| Aneuploidy score ()5 |  |  |
| Heterogeneity score ()5 |  |  |
| Ploidy proportion\* |  |  |

**Table 1. Scores and indices used to quantify aneuploidy.** For the instability index, = the number of unique chromosomes examined. , where is the number of cells containing the modal number of the chromosome. is the number of cells examined. For the ANCA score, is the number of non-diploid chromosomes observed in the cell. For the aneuploidy score, is the copy number of the cell at the chromosome. is the euploid copy number at the chromosome. For the heterogeneity score, is the total number of copy number states. is the number of cells with copy number state at bin . is ordered such that . For the ploidy proportion, is the proportion of cells that are aneuploid, is the proportion of cells that are diploid, and is the proportion of cells that are polyploid.

**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 are 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. However,

*Approaches to detecting numerical chromosomal variation*

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.

*Study design for investigating mechanisms of aneuploidy*

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.

*The importance of characterizing ploidy*

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 autosome in a cell, and is the physiologic state of most non-cancerous human cells. Polyploidy refers to an integer-valued increase in the number of chromosomes, often resulting from whole-genome duplication. Polyploidization is a characteristic of cancer cells that resume growth after chemotherapeutic treatment *in vitro*6 and is thus important to study. Aneuploidy occurs when the copy number of 1 or more chromosomes differs from the others and is a feature of many cancers.

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. Aneuvis is freely available for researchers to use online, and the source code is available online.

**Results**

*Overview of input*

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). Aneuvis sits downstream of 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

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**).

*Input*

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 Ginkgo7, 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.

*Overview of output*

The output from aneuvis is organized into 2 sections – table summary and visualization – that are generated automatically from the input data. The table summary includes statistics that summarize the degree of numerical chromosomal aneuploidy and heterogeneity in a population of single cells. The visualization tab includes plots derived from the statistics available in table summary. These visualizations enable comparisons between groups and between platforms. In order to test statistical significance between groups, we also incorporate automatic permutation testing between groups along with a visualization of the output. In sum, the output from aneuvis provides a numerical and visual summary of the ploidy state of populations of single cells across different treatments and experimental platforms. This output has the potential to uncover unexpected relationships in terms of the aneuploid state between different treatments, which might hint at a similar underlying mechanism. This platform will empower researchers with a new, automated approach to view aneuploidy between groups.

*Table summary*

From searching the literature, we identified several statistics that can be used to summarize the chromosomal state across entire populations of cells or across individual chromosomes. Table summary is divided into two parts – aggregate and chromosome-level summaries per group. The aggregate summary contains 5 distinct statistics that were defined in previous studies (**Table 1**). The same statistical are calculated at the chromosomal level, in order to identify those chromosomes that are most perturbed within each group. The chromosome-level summaries are available for the same statistics, with the exception of the ploidy proportion, which is only calculated as the proportion of cells that are diploid. The user interface to the summary table is dynamic and searchable, and the data is downloadable. Below is a description of each of the scores.

The instability index (I) is a metric that calculates the percentage of cells that contain a chromosomal aberration. This metric does not directly depend on the number of chromosomes; however, measuring more chromosomes may make it more likely to detect at least one chromosome that contains an abnormal number of copies.

The Average Number of Copy Number Alterations (ANCA) score has been applied in the context of colorectal and cervical cancer in an attempt to quantify the relationship between tumor aggressiveness and genomic instability3,4. Previous studies have uncovered that more aggressive tumors had a higher ANCA score. However, one limitation of the ANCA score is that it does not account for the number of chromosomes examined. Within aneuvis, we introduce a derivative of the ANCA score, called the Normalized ANCA score, which accounts for the number of chromosomes measured and enables comparisons of this metric between experiments that utilize different numbers of probes.

The aneuploidy (D) and heterogeneity (H) scores were derived from Bakker et al and represent a pair of statistics that account for the number of cells and chromosomes tested for. The aneuploidy score increases with an increased chromosome copy number – the only score to take the number of chromosomes into account. The heterogeneity score increases with the number of distinct chromosomal states observed, and is maximized when each cell has a distinct state. In contrast to the aneuploidy score, the heterogeneity score does not incorporate the chromosomal copy number. These statistics were derived for looking at copy number data from whole genome single cell sequencing data, though their formulation enables them to be applied to other copy number datasets.

*Visualization*

Graphical outputs are automatically generated and are divided into visualizations that are data-type specific, and those that are integrated between different datatypes. For example, we introduce a bivariate percentage plot for FISH data, which is often performed with probes that can quantify the ploidy of between 2 and 4 chromosomes.

*Integrated analyses*

The aneuploidy and heterogeneity scores are generated for each group as a whole, as well as for each individual chromosome, as a scatterplot. In addition, we display the ploidy proportion data across all samples using a ternary plot, which is useful for representing 3 variables constrained to sum to 1. The ternary plot includes groups from all platforms, and enables comparisons between these groups. The

*Permutation*

Oftentimes, a research question may involve asking whether two treatment groups are different from each other in terms of their degree of chromosomal instability. Permutation testing is a robust approach that harnesses computational resources to generate statistical conclusions using an intuitive, assumption-free approach. Here, we implement permutation testing between all pairwise comparisons by shuffling the labels associated with each observed cell. We test the null hypothesis that each group has an equal degree of chromosomal instability relative to all other groups, and reject the null at a p value of 0.05.



**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.

**Discussion**

Chromosomal instability refers to a biological phenomenon whereby an actively dividing population of cells fails to maintain a consistent number of chromosomes after cell division.

The presence of WGD alters sensitivity of cells to potential cancer therapeutics and is thus important to detect8,9.

In addition, polypoid cells are thought to be less sensitive to the loss or gain of chromosomes than diploid cells, as their genome is buffered with the extra set of chromosomes10.

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.

**Conclusion and Future directions**

Future directions include adding statistical and visualization support for more complex study designs, such as nested designs and time series experiments.

**Materials and Methods**

Aneuvis was created using Shiny version 1.0.5 (R version 3.4.3) and is available under a GPLv3 license. Permutation testing is set to 250 permutations by default. Source code and versioning history is available on github.

**Bibliography**

1. Lengauer, C., Kinzler, K. W. & Vogelstein, B. Genetic instability in colorectal cancers. *Nature* **386,** 623–627 (1997).

2. Bayani, J. *et al.* Distinct Patterns of Structural and Numerical Chromosomal Instability Characterize Sporadic Ovarian Cancer. *Neoplasia* **10,** 1057–IN3 (2008).

3. Blegen, H. *et al.* DNA amplifications and aneuploidy, high proliferative activity and impaired cell cycle control characterize breast carcinomas with poor prognosis. *Anal. Cell. Pathol.* **25,** 103–14 (2003).

4. Ried, T., Heselmeyer-Haddad, K., Blegen, H., Schröck, E. & Auer, G. Genomic changes defining the genesis, progression, and malignancy potential in solid human tumors: A phenotype/genotype correlation. *Genes Chromosom. Cancer* **25,** 195–204 (1999).

5. Bakker, B. *et al.* Single-cell sequencing reveals karyotype heterogeneity in murine and human malignancies. *Genome Biol.* **17,** 115 (2016).

6. Mosieniak, G. *et al.* Polyploidy Formation in Doxorubicin-Treated Cancer Cells Can Favor Escape from Senescence. *Neoplasia* **17,** 882–893 (2015).

7. Garvin, T. *et al.* Interactive analysis and assessment of single-cell copy-number variations. *Nat. Methods* **12,** 1058–1060 (2015).

8. Mason, J. M. *et al.* Functional characterization of CFI-402257, a potent and selective Mps1/TTK kinase inhibitor, for the treatment of cancer. *Proc. Natl. Acad. Sci. U. S. A.* **114,** 3127–3132 (2017).

9. Jemaà, M. *et al.* Whole-genome duplication increases tumor cell sensitivity to MPS1 inhibition. *Oncotarget* **7,** 885–901 (2016).

10. Rajagopalan, H., Nowak, M. A., Vogelstein, B. & Lengauer, C. The significance of unstable chromosomes in colorectal cancer. *Nat. Rev. Cancer* **3,** 695–701 (2003).

11. Rösl, F., Arab, A., Klevenz, B. & Hausen, H. The effect of DNA methylation on gene regulation of human papillomaviruses. *J. Gen. Virol.* 791–801 (1993).