**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**

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. The automated analysis of these datatypes 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 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.

**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.

*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.

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.

*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*1 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.

The presence of WGD alters sensitivity of cells to potential cancer therapeutics and is thus important to detect2,3.

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 chromosomes4.

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.

When viewed as a population, the

**Results**

*Input*

The application accepts three different types of input. First, bed files that include copy number output from Ginkgo5, a web-based program for calling ploidy changes from whole-genome sequencing data, may be uploaded. In addition, multicolor fluorescent in situ hybridization (FISH) data summarized at the chromosomal level can be uploaded as a Microsoft Excel file. An approach using interphase FISH to estimate chromosomal copy number in various tissues was recently described by one of us6. This data is often recorded into excel spreadsheets, and the file structure accepted by aneuvis are described on the website. Lastly, spectral karyotyping (SKY) data from a Microsoft Excel spreadsheet can be accepted into Aneuvis. Each of these data types can be characterized by a cell-by-chromosome matrix that contains integer values representing the copy number state of each chromosome.

*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. First, the ANCA score has been reported previously7,8. 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.

generated from previous outputs summaries per chromosome and summaries by group.

*Processing*

Others have described an instability index, which is essentially identical to the

|  |  |  |  |
| --- | --- | --- | --- |
|  | Reference | Calculation | Support |
| Instability index () | 9,10 Lengauer et al., 1997; Bayani et al. 2008 | = Number of unique chromosomes examined  , where is the number of cells containing the modal number of the chromosome. is the number of cells examined. |  |
| ANCA () – Average # of CN alterations | 7,8 Ried *et al.* 1999; Blegen *et al*., 2003 | is the number of chromosomal aberrations observed in the cell. |  |
| Normalized ANCA () | None | is the number of chromosomal aberrations observed in the cell. |  |
| Aneuploidy score () | 11 Bakker et al, 2016 | is the copy number of the cell at the chromosome. is the euploid copy number at the chromosome. |  |
| Heterogeneity score () | 11 Bakker et al, 2016 | is the total number of copy number states. is the number of cells with copy number state at bin . |  |

Table 1. Scores and indices used to quantify aneuploidy.

We are the first to provide a user-friendly, permutation-based approach for calculating statistical significance between different groups.

We provide a standardized table, complete with potential uses and the interpretation of each statistic.

|  |  |  |  |
| --- | --- | --- | --- |
| Method | Chromosome-level summary | Group-level summary | Interpretation |
| Instability index () | Yes | Yes | A measure of the number of cells that are maintain . |
| ANCA () – Average # of CN alterations | Yes | Yes | The average number of copy number alterations within a set of related samples (ie cells from the same tumor) |
| Normalized ANCA () | Yes | Yes | The average number of copy number alterations within a set of related samples (ie cells from the same tumor), normalized by the number of chromosomes probed for. |
| Aneuploidy score () | Yes | Yes |  |
| Heterogeneity score () | Yes | Yes |  |
| Percent ploidy | - | Yes |  |

appropriate uses and interpretation

Visualization

Permutation

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*

There are two distinct approaches to measuring aneuploidy – single cell and multi-cell approaches. Single cell approaches are most appropriate for analyzing variability and for assessing chromosomal instability in a population of single cells.

“Appropriate statistical means should be used to establish if the rate of chromosomal changes in a test population differs significantly from a reference population.” -

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

**Bibliography**

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

2. 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).

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

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

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

6. Andriani, G. A. & Montagna, C. FISH Analysis of Aging-Associated Aneuploidy in Neurons and Nonneuronal Brain Cells. in *Genomic Mosaicism in Neurons and Other Cell Types* 271–298 (2017). doi:10.1007/978-1-4939-7280-7

7. 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).

8. 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).

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

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

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