

that corresponds to a subset of at least one target nucleic acid in the sample and a second portion that encodes a respective identifier (e.g., bar code) for the respective sequencing read in a plurality of identifiers. Each respective identifier is independent of the sequence of the at least one target nucleic acid. The plurality of sequencing reads collectively includes the plurality of identifiers. A visualization tool is displayed. A request is obtained from a user through the visualization tool. The request specifies a genomic region represented by the nucleic acid sequencing dataset. Responsive to obtaining the request, the request is parsed by obtaining a plurality of sequencing reads within the genomic region from the nucleic acid sequencing dataset. A scan window is run against the plurality of sequencing reads thereby creating a plurality of windows, each respective window of the plurality of windows corresponding to a different region of the genomic region and including an identity of each identifier of each sequencing read in the different region of the genomic region in the nucleic acid sequencing dataset. A two dimensional heat map that represents each possible window pair in the plurality of windows is displayed. Each respective window pair is displayed in the two dimensional heat map as a color selected from a color scheme based upon the number of identifiers in common in the respective window pair.

[0026] Various embodiments of systems, methods and devices within the scope of the appended claims each have several aspects, no single one of which is solely responsible for the desirable attributes described herein. Without limiting the scope of the appended claims, some prominent features are described herein. After considering this discussion, and particularly after reading the section entitled “Detailed Description” one will understand how the features of various embodiments are used.

[0026A] In accordance with the above aspects, a first implementation provides a system for providing structural variation or phasing information over a network connection to a remote client computer, the system comprising one or more microprocessors, a persistent memory and a non-persistent memory, the persistent memory and the non-persistent memory collectively storing one or more nucleic acid sequence datasets, wherein each respective nucleic acid sequencing dataset in the one or more nucleic acid sequence datasets

corresponds to at least one target nucleic acid in a respective sample in a plurality of samples, wherein the respective sample is associated with a genome of at least one species, the respective nucleic acid sequencing dataset comprises (i) a header, (ii) a synopsis, and (iii) a data section, the data section comprises a plurality of sequencing reads, each respective sequencing read in the plurality of sequencing reads comprises a first portion that corresponds to a subset of at least one target nucleic acid in the respective sample and a second portion that encodes a respective identifier for the respective sequencing read in a plurality of identifiers, each respective identifier is independent of the sequence of the at least one target nucleic acid, and the plurality of sequencing reads collectively includes the plurality of identifiers, and wherein the persistent memory and the non-persistent memory further collectively store one or more programs that use the one or more microprocessors to: provide a visualization tool for installation on the remote client computer; obtain a request, sent from the remote client computer from a user, over a network connection, for structural variation or phasing information using a first dataset in the one or more datasets, wherein the request is formatted in accordance with an expression syntax; and responsive to obtaining the request, automatically parse the request by: i) loading the header and the synopsis of the first dataset into the non-persistent memory if not already loaded into the non-persistent memory while retaining the data section in persistent memory, (ii) comparing the request to the synopsis of the first dataset thereby identifying one or more portions of the data section of the first dataset, (iii) loading the one or more identified portions of the data section into non-persistent memory, wherein the loading loads less than the entirety of the data section, (iv) formatting structural variation or phasing information for display on the client computer using the first dataset, and (v) transmitting the formatted structural variation or phasing information over the network connection to the remote client computer for display on the remote client computer.

[0026B] A second implementation provides a system for providing structural variation or phasing information, the system comprising one or more microprocessors, a persistent memory and a non-persistent memory, the persistent memory and the non-persistent

memory collectively storing one or more nucleic acid sequence datasets, wherein each respective nucleic acid sequencing dataset in the one or more nucleic acid sequence datasets corresponds to at least one target nucleic acid in a respective sample in a plurality of samples, wherein the respective sample is associated with a genome of at least one species, the respective nucleic acid sequencing dataset comprises (i) a header, (ii) a synopsis, and (iii) a data section, the data section comprises a plurality of sequencing reads, each respective sequencing read in the plurality of sequencing reads comprises a first portion that corresponds to a subset of at least one target nucleic acid in the respective sample and a second portion that encodes a respective identifier for the respective sequencing read in a plurality of identifiers, each respective identifier is independent of the sequence of the at least one target nucleic acid, and the plurality of sequencing reads collectively includes the plurality of identifiers, and wherein the persistent memory and the non-persistent memory further collectively store one or more programs that use the one or more microprocessors to: provide a visualization tool; obtain a request from a user, through the visualization tool, for structural variation or phasing information using a first dataset in the one or more datasets, wherein the request is formatted in accordance with an expression syntax, responsive to obtaining the request, automatically parse the request by: (i) loading the header and the synopsis of the first dataset into the non-persistent memory if not already loaded into the non-persistent memory while retaining the data section in persistent memory, (ii) comparing the request to the synopsis of the first dataset thereby identifying one or more portions of the data section of the first dataset, (iii) loading the one or more identified portions of the data section into non- persistent memory, wherein the loading loads less than the entirety of the data section, (iv) formatting structural variation or phasing information for display in the visualization tool using the first dataset, and (v) displaying the formatted structural variation or phasing information in the visualization tool.

[0026C] A third implementation provides a system for obtaining structural variation or phasing information over a network connection from a remote computer, wherein the system comprises one or more microprocessors, and a memory that stores one or more programs, wherein the one or more programs use the one or more microprocessors to

execute a method comprising: (A) invoking a visualization tool; (B) obtaining, through the visualization tool from a user, a request for structural variation or phasing information in a first nucleic acid sequencing dataset from among one or more nucleic acid sequencing datasets stored on the remote computer, wherein each respective nucleic acid sequencing dataset in the one or more nucleic acid sequence datasets corresponds to at least one target nucleic acid in a respective sample in a plurality of samples, wherein the request is formatted in accordance with an expression syntax, the respective sample is associated with a genome of at least one species, the respective nucleic acid sequencing dataset comprises (i) a header, (ii) a synopsis, and (iii) a data section, the data section comprises a plurality of sequencing reads, each respective sequencing read in the plurality of sequencing reads comprises a first portion that corresponds to a subset of at least one target nucleic acid in the respective sample and a second portion that encodes a respective identifier for the respective sequencing read in a plurality of identifiers, each respective identifier is independent of the sequence of the at least one target nucleic acid, and the plurality of sequencing reads collectively includes the plurality of identifiers; (C) sending the request to the remote computer over the network connection, wherein the remote computer has persistent memory and non-persistent memory, thereby causing the remote computer to execute a method comprising: (i) loading the header and the synopsis of the first dataset into the non-persistent memory if not already loaded into the non-persistent memory of the remote computer while retaining the data section in persistent memory, (ii) comparing the request to the synopsis of the first dataset thereby identifying one or more portions of the data section of the first dataset, (iii) loading the one or more identified portions of the data section into non- persistent memory, wherein the loading loads less than the entirety of the data section, and (iv) formatting structural variation or phasing information; and (D) receiving the formatted structural variation or phasing information over the network connection from the remote computer for display in the visualization tool.

[0026D] A fourth implementation provides a system for providing structural variation or phasing information, the system comprising one or more microprocessors, and a memory, wherein the system has access to one or more nucleic acid sequence datasets,

wherein each respective nucleic acid sequencing dataset in the one or more nucleic acid sequence datasets corresponds to at least one target nucleic acid in a respective sample in a plurality of samples, wherein the respective sample is associated with a genome of at least one species, the respective nucleic acid sequencing dataset comprises (i) a header, (ii) a synopsis, and (iii) a data section, the data section comprises a plurality of sequencing reads, each respective sequencing read in the plurality of sequencing reads comprises a first portion that corresponds to a subset of at least one target nucleic acid in the respective sample and a second portion that encodes a respective identifier for the respective sequencing read in a plurality of identifiers, each respective identifier is independent of the sequence of the at least one target nucleic acid, and the plurality of sequencing reads collectively includes the plurality of identifiers, and wherein the memory further collectively stores one or more programs that use the one or more microprocessors to: provide a visualization tool; obtain a request from a user, through the visualization tool, for structural variation or phasing information using a first dataset in the one or more datasets, wherein the request is formatted in accordance with an expression syntax, responsive to the obtained request, automatically parse the request by: (i) comparing the request to the synopsis of the first dataset thereby identifying one or more portions of the data section of the first dataset, (ii) formatting structural variation or phasing information for display in the visualization tool using the identified one or more portions of the data section of the first dataset, wherein the one or more portions of the data section of the first dataset is less than the entirety of the data section of the first dataset, and (iii) displaying the formatted structural variation or phasing information in the visualization tool.

[0026E] A fifth implementation provides a system for providing structural variation or phasing information over a network connection to a remote client computer, the system comprising one or more microprocessors, a persistent memory and a non-persistent memory, the persistent memory and the non-persistent memory collectively storing one or more nucleic acid sequence datasets, wherein each respective nucleic acid sequencing dataset in the one or more nucleic acid sequence datasets corresponds to at least one target nucleic acid in a respective sample in a plurality of samples, wherein the respective sample

is associated with a genome of at least one species, the respective nucleic acid sequencing dataset comprises (i) a header, (ii) a synopsis, and (iii) a data section, wherein the header delineates a plurality of components in the respective nucleic acid sequencing dataset and wherein the header further comprises the version of the dataset structure used by the nucleic acid sequencing dataset, the data section comprises a plurality of sequencing reads, each respective sequencing read in the plurality of sequencing reads comprises a first portion that corresponds to a subset of at least one target nucleic acid in the respective sample and a second portion that encodes a respective identifier for the respective sequencing read in a plurality of identifiers, each respective identifier is independent of the sequence of the at least one target nucleic acid, and the plurality of sequencing reads collectively includes the plurality of identifiers, and wherein the persistent memory and the non-persistent memory further collectively store one or more programs that use the one or more microprocessors to: provide a visualization tool for installation on the remote client computer; obtain a request, sent from the remote client computer from a user, over a network connection, for structural variation or phasing information using a first dataset in the one or more datasets, and responsive to obtaining the request, automatically parse the request by: (i) loading the header and the synopsis of the first dataset into the non-persistent memory if not already loaded into the non-persistent memory while retaining the data section in persistent memory, (ii) comparing the request to the synopsis of the first dataset thereby identifying one or more portions of the data section of the first dataset, (iii) loading the one or more identified portions of the data section into non-persistent memory, wherein the loading loads less than the entirety of the data section, (iv) formatting structural variation or phasing information for display on the client computer using the first dataset, and (v) transmitting the formatted structural variation or phasing information over the network connection to the remote client computer for display on the remote client computer.

[0026F] A sixth implementation provides a system for providing structural variation or phasing information, the system comprising one or more microprocessors, a persistent memory and a non-persistent memory, the persistent memory and the non-persistent memory collectively storing one or more nucleic acid sequence datasets, wherein each

respective nucleic acid sequencing dataset in the one or more nucleic acid sequence datasets corresponds to at least one target nucleic acid in a respective sample in a plurality of samples, wherein the respective sample is associated with a genome of at least one species, the respective nucleic acid sequencing dataset comprises (i) a header, (ii) a synopsis, and (iii) a data section wherein the header delineates a plurality of components in the respective nucleic acid sequencing dataset wherein the plurality of components comprises an index to read data wherein the index to read data comprises a lookup table between a respective identifier in the plurality of identifiers and a shortened version of the respective identifier and wherein the plurality of components further comprises at least one of a summary, an index to variant call data, a phase block track, a refseq index track, a gene track, an exon track, a structural variant dataset track, an index to a target dataset, and an index to a fragment dataset, the data section comprises a plurality of sequencing reads, each respective sequencing read in the plurality of sequencing reads comprises a first portion that corresponds to a subset of at least one target nucleic acid in the respective sample and a second portion that encodes a respective identifier for the respective sequencing read in a plurality of identifiers, each respective identifier is independent of the sequence of the at least one target nucleic acid, and the plurality of sequencing reads collectively includes the plurality of identifiers, and wherein the persistent memory and the non-persistent memory further collectively store one or more programs that use the one or more microprocessors to: provide a visualization tool; obtain a request from a user, through the visualization tool, for structural variation or phasing information using a first dataset in the one or more datasets, responsive to obtaining the request, automatically parse the request by: (i) loading the header and the synopsis of the first dataset into the non-persistent memory if not already loaded into the non-persistent memory while retaining the data section in persistent memory, (ii) comparing the request to the synopsis of the first dataset thereby identifying one or more portions of the data section of the first dataset, (iii) loading the one or more identified portions of the data section into non- persistent memory, wherein the loading loads less than the entirety of the data section, (iv) formatting structural variation or phasing information for display in the visualization tool using the first dataset, and (v) displaying the

formatted structural variation or phasing information in the visualization tool.

[0026G] A seventh implementation provides a system for obtaining phasing information over a network connection from a remote computer, wherein the system comprises one or more microprocessors, and a memory that stores one or more programs, wherein the one or more programs use the one or more microprocessors to execute a method comprising: (A) invoking a visualization tool; (B) obtaining, through the visualization tool from a user, a request for phasing information in a first nucleic acid sequencing dataset from among one or more nucleic acid sequencing datasets stored on the remote computer, wherein each respective nucleic acid sequencing dataset in the one or more nucleic acid sequence datasets corresponds to at least one target nucleic acid in a respective sample in a plurality of samples, wherein the respective sample is associated with a genome of at least one species, the respective nucleic acid sequencing dataset comprises (i) a header, (ii) a synopsis, and (iii) a data section, the data section comprises a plurality of sequencing reads, each respective sequencing read in the plurality of sequencing reads comprises a first portion that corresponds to a subset of at least one target nucleic acid in the respective sample and a second portion that encodes a respective identifier for the respective sequencing read in a plurality of identifiers, each respective identifier is independent of the sequence of the at least one target nucleic acid, and the plurality of sequencing reads collectively includes the plurality of identifiers; (C) sending the request to the remote computer over the network connection, wherein the remote computer has persistent memory and non-persistent memory, thereby causing the remote computer to execute a method comprising: (i) loading the header and the synopsis of the first dataset into the non-persistent memory if not already loaded into the non-persistent memory of the remote computer while retaining the data section in persistent memory, (ii) comparing the request for phasing information to the synopsis of the first dataset thereby identifying one or more portions of the data section of the first dataset, (iii) loading the one or more identified portions of the data section into non- persistent memory, wherein the loading loads less than the entirety of the data section, and (iv) formatting phasing information; and (D) receiving the formatted phasing information over

the network connection from the remote computer for display in the visualization tool, herein the request is for phasing information in a region of the genome and the formatted phasing information includes a graphic representation comprising: a first haplotype track corresponding to a first parental haplotype of a first species in the at least one species in the region of the genome for the first dataset, a second haplotype track, corresponding to a second parental haplotype of the first species in the region of the genome for the first dataset, and an indeterminate track corresponding to regions of the at least one nucleic acid sample that have not been assigned a parental haplotype in the region of the genome for the first dataset.

[0026H] An eighth implementation provides a system for providing structural variation or phasing information over a network connection to a remote client computer, the system comprising one or more microprocessors, a persistent memory and a non-persistent memory, the persistent memory and the non-persistent memory collectively storing one or more nucleic acid sequence datasets, wherein each respective nucleic acid sequencing dataset in the one or more nucleic acid sequence datasets corresponds to at least one target nucleic acid in a respective sample in a plurality of samples, wherein the respective sample is associated with a genome of a plurality of species and includes at least a portion of the genome of a first species and a portion of the genome of the second species and, wherein the respective sample is associated with a genome of at least one species, the respective nucleic acid sequencing dataset comprises (i) a header, (ii) a synopsis, and (iii) a data section, the data section comprises a plurality of sequencing reads, each respective sequencing read in the plurality of sequencing reads comprises a first portion that corresponds to a subset of at least one target nucleic acid in the respective sample and a second portion that encodes a respective identifier for the respective sequencing read in a plurality of identifiers, each respective identifier is independent of the sequence of the at least one target nucleic acid, and the plurality of sequencing reads collectively includes the plurality of identifiers, and wherein the persistent memory and the non-persistent memory further collectively store one or more programs that use the one or more microprocessors to: provide a visualization tool for installation on the

remote client computer; obtain a request, sent from the remote client computer from a user, over a network connection, for structural variation or phasing information using a first dataset in the one or more datasets; and responsive to obtaining the request, automatically parse the request by: (i) loading the header and the synopsis of the first dataset into the non- persistent memory if not already loaded into the non-persistent memory while retaining the data section in persistent memory, (ii) comparing the request to the synopsis of the first dataset thereby identifying one or more portions of the data section of the first dataset, (iii) loading the one or more identified portions of the data section into non-persistent memory, wherein the loading loads less than the entirety of the data section, (iv) formatting structural variation or phasing information for display on the client computer using the first dataset, and (v) transmitting the formatted structural variation or phasing information over the network connection to the remote client computer for display on the remote client computer.

[0026I] A ninth implementation provides a system for processing program output over a network connection using a local computer, wherein the local computer comprises one or more microprocessors, and a memory that stores one or more programs, wherein the one or more programs use the one or more microprocessors to execute a method in accordance with a first operating system running on the local computer, the method comprising: (A) invoking a first instance of a first program; (B) obtaining, through the first instance of the first program from a user, a login and a password to a user account on a remote computer; (C) logging the user into the user account on the remote computer automatically, using the login and the password provided by the first instance of the first program, across a network connection between the local computer and the remote computer; (D) responsive to successful login on the remote computer, automatically sending, without human intervention, a second instance of the first program configured to auto-install on the remote computer upon transmission to the remote computer; (E) receiving from the remote computer a request to open a panel within the first instance of the first program, wherein the panel is originated by the second instance of the first program running on the remote computer, and wherein the panel solicits input from the user for controlling the second instance of the first program; (F) responsive to receiving

input from the user for controlling the second instance of the first program in the panel on the local computer, sending the input to the second instance of the first program on the remote computer; and (G) receiving, from the remote computer across the network connection, output from the second instance of the first program responsive to the input and displaying the output at the local computer.

[0026J] A tenth implementation provides a system for viewing nucleic acid sequencing data, the system comprising one or more microprocessors and a memory, wherein the memory stores one or more programs that use the one or more microprocessors to: obtain a nucleic acid sequencing dataset corresponding to at least one target nucleic acid in a sample, the nucleic acid sequencing dataset comprising a plurality of sequencing reads from the sample, wherein each respective sequencing read in the plurality of sequencing reads comprises a first portion that corresponds to a subset of at least one target nucleic acid in the sample and a second portion that encodes a respective identifier for the respective sequencing read in a plurality of identifiers, each respective identifier is independent of the sequence of the at least one target nucleic acid, and the plurality of sequencing reads collectively includes the plurality of identifiers, display a visualization tool; obtain a request from a user, through the visualization tool, that specifies a genomic region represented by the nucleic acid sequencing dataset; and responsive to the obtained request, parse the request by: (i) obtaining a plurality of sequencing reads within the genomic region from the nucleic acid sequencing dataset, (ii) running a scan window against the plurality of sequencing reads thereby creating a plurality of windows, each respective window of the plurality of windows corresponding to a different region of the genomic region and including an identity of each identifier of each sequencing read in the different region of the genomic region in the nucleic acid sequencing dataset, and (iii) displaying a two dimensional heat map that represents each possible window pair in the plurality of windows, wherein each respective window pair is displayed in the two dimensional heat map as a color selected from a color scheme based upon the number of identifiers in common in the respective window pair, wherein the visualization tool provides a zoom affordance to zoom the heat map or a pan affordance to pan the heat map.

[0026K] An eleventh implementation provides a method for providing structural variation or phasing information over a network connection to a remote client computer, the method comprising: at a system comprising one or more microprocessors, a persistent memory and a non-persistent memory, the persistent memory and the non-persistent memory collectively storing one or more nucleic acid sequence datasets, wherein each respective nucleic acid sequencing dataset in the one or more nucleic acid sequence datasets corresponds to at least one target nucleic acid in a respective sample in a plurality of samples, wherein the respective sample is associated with a genome of at least one species, the respective nucleic acid sequencing dataset comprises (i) a header, (ii) a synopsis, and (iii) a data section, the data section comprises a plurality of sequencing reads, each respective sequencing read in the plurality of sequencing reads comprises a first portion that corresponds to a subset of at least one target nucleic acid in the respective sample and a second portion that encodes a respective identifier for the respective sequencing read in a plurality of identifiers, each respective identifier is independent of the sequence of the at least one target nucleic acid, and the plurality of sequencing reads collectively includes the plurality of identifiers: (A) providing a visualization tool for installation on the remote client computer; (B) obtaining a request, sent from the remote client computer from a user, over a network connection, for structural variation or phasing information using a first dataset in the one or more datasets, wherein the request is formatted in accordance with an expression syntax; and (C) responsive to obtaining the request, automatically parsing the request by: (i) loading the header and the synopsis of the first dataset into the non-persistent memory if not already loaded into the non-persistent memory while retaining the data section in persistent memory, (ii) comparing the request to the synopsis of the first dataset thereby identifying one or more portions of the data section of the first dataset, (iii) loading the one or more identified portions of the data section into non-persistent memory, wherein the loading loads less than the entirety of the data section, (iv) formatting structural variation or phasing information for display on the client computer using the first dataset, and (v) transmitting the formatted structural variation or phasing information over the network connection to the remote client computer for display on the

remote client computer.

[0026L] A twelfth implementation provides a method for providing structural variation or phasing information, the method comprising: at a system comprising one or more microprocessors, a persistent memory and a non-persistent memory, the persistent memory and the non-persistent memory collectively storing one or more nucleic acid sequence datasets, wherein each respective nucleic acid sequencing dataset in the one or more nucleic acid sequence datasets corresponds to at least one target nucleic acid in a respective sample in a plurality of samples, wherein the respective sample is associated with a genome of at least one species, the respective nucleic acid sequencing dataset comprises (i) a header, (ii) a synopsis, and (iii) a data section, the data section comprises a plurality of sequencing reads, each respective sequencing read in the plurality of sequencing reads comprises a first portion that corresponds to a subset of at least one target nucleic acid in the respective sample and a second portion that encodes a respective identifier for the respective sequencing read in a plurality of identifiers, each respective identifier is independent of the sequence of the at least one target nucleic acid, and the plurality of sequencing reads collectively includes the plurality of identifiers: (A) providing a visualization tool; (B) obtaining a request from a user, through the visualization tool, for structural variation or phasing information using a first dataset in the one or more datasets wherein the request is formatted in accordance with an expression syntax, (C) responsive to obtaining the request, automatically parsing the request by: (i) loading the header and the synopsis of the first dataset into the non-persistent memory if not already loaded into the non-persistent memory while retaining the data section in persistent memory, (ii) comparing the request to the synopsis of the first dataset thereby identifying one or more portions of the data section of the first dataset, (iii) loading the one or more identified portions of the data section into non-persistent memory, wherein the loading loads less than the entirety of the data section, (iv) formatting structural variation or phasing information for display in the visualization tool using the first dataset, and (v) displaying the formatted structural variation or phasing information in the visualization tool.

[0026M] A thirteenth implementation provides a method for obtaining structural

variation or phasing information over a network connection from a remote computer, the method comprising: at a system comprising one or more microprocessors, and a memory that stores one or more programs: (A) invoking a visualization tool; (B) obtaining, through the visualization tool from a user, a request for structural variation or phasing information in a first nucleic acid sequencing dataset from among one or more nucleic acid sequencing datasets stored on the remote computer, wherein each respective nucleic acid sequencing dataset in the one or more nucleic acid sequence datasets corresponds to at least one target nucleic acid in a respective sample in a plurality of samples, wherein the request is formatted in accordance with an expression syntax, the respective sample is associated with a genome of at least one species, the respective nucleic acid sequencing dataset comprises (i) a header, (ii) a synopsis, and (iii) a data section, the data section comprises a plurality of sequencing reads, each respective sequencing read in the plurality of sequencing reads comprises a first portion that corresponds to a subset of at least one target nucleic acid in the respective sample and a second portion that encodes a respective identifier for the respective sequencing read in a plurality of identifiers, each respective identifier is independent of the sequence of the at least one target nucleic acid, and the plurality of sequencing reads collectively includes the plurality of identifiers; (C) sending the request to the remote computer over the network connection, wherein the remote computer has persistent memory and non-persistent memory, thereby causing the remote computer to execute a method comprising: (i) loading the header and the synopsis of the first dataset into the non-persistent memory if not already loaded into the non-persistent memory of the remote computer while retaining the data section in persistent memory, (ii) comparing the request to the synopsis of the first dataset thereby identifying one or more portions of the data section of the first dataset, (iii) loading the one or more identified portions of the data section into non- persistent memory, wherein the loading loads less than the entirety of the data section, and (iv) formatting structural variation or phasing information; and (D) receiving the formatted structural variation or phasing information over the network connection from the remote computer for display in the visualization tool.

[0026N] A fourteenth implementation provides a method for providing structural variation or phasing information, the method comprising: at a system comprising one or more microprocessors, and a memory, wherein the system has access to one or more nucleic acid sequence datasets, wherein each respective nucleic acid sequencing dataset in the one or more nucleic acid sequence datasets corresponds to at least one target nucleic acid in a respective sample in a plurality of samples, wherein the respective sample is associated with a genome of at least one species, the respective nucleic acid sequencing dataset comprises (i) a header, (ii) a synopsis, and (iii) a data section, the data section comprises a plurality of sequencing reads, each respective sequencing read in the plurality of sequencing reads comprises a first portion that corresponds to a subset of at least one target nucleic acid in the respective sample and a second portion that encodes a respective identifier for the respective sequencing read in a plurality of identifiers, each respective identifier is independent of the sequence of the at least one target nucleic acid, and the plurality of sequencing reads collectively includes the plurality of identifiers, (A) providing a visualization tool; (B) obtaining a request from a user, through the visualization tool, for structural variation or phasing information using a first dataset in the one or more datasets, wherein the request is formatted in accordance with an expression syntax; and (C) responsive to obtaining the request, automatically parsing the request by a method comprising: (i) comparing the request to the synopsis of the first dataset thereby identifying one or more portions of the data section of the first dataset, (ii) formatting structural variation or phasing information for display in the visualization tool using the identified one or more portions of the data section of the first dataset, wherein the one or more portions of the data section of the first dataset is less than the entirety of the data section of the first dataset, and (iii) displaying the formatted structural variation or phasing information in the visualization tool.

[0026O] A fifteenth implementation provides a method for processing program output over a network connection, the method comprising: at a local computer, wherein the local computer comprises one or more microprocessors, and a memory that stores one or more programs to: (A) invoke a first instance of a first program; (B) obtain, through the first

instance of the first program from a user, a login and a password to a user account on a remote computer; (C) log the user into the user account on the remote computer automatically, using the login and the password provided by the first instance of the first program, across a network connection between the local computer and the remote computer; (D) responsive to successful login on the remote computer, automatically send, without human intervention, a second instance of the first program configured to auto-install on the remote computer upon transmission to the remote computer; (E) receive from the remote computer a request to open a panel within the first instance of the first program, wherein the panel is originated by the second instance of the first program running on the remote computer, and wherein the panel solicits input from the user for controlling the second instance of the first program; (F) responsive to receiving input from the user for controlling the second instance of the first program in the panel on the local computer, send the input to the second instance of the first program on the remote computer; and (G) receive, from the remote computer across the network connection, output from the second instance of the first program responsive to the input and displaying the output at the local computer.

[0026P] A sixteenth implementation provides a method for viewing nucleic acid sequencing data, the method comprising: at a system comprising one or more microprocessors and a memory, wherein the memory stores one or more programs that use the one or more microprocessors to: obtain a nucleic acid sequencing dataset corresponding to at least one target nucleic acid in a sample, the nucleic acid sequencing dataset comprising a plurality of sequencing reads from the sample, wherein each respective sequencing read in the plurality of sequencing reads comprises a first portion that corresponds to a subset of at least one target nucleic acid in the sample and a second portion that encodes a respective identifier for the respective sequencing read in a plurality of identifiers, each respective identifier is independent of the sequence of the at least one target nucleic acid, and the plurality of sequencing reads collectively includes the plurality of identifiers, display a visualization tool; obtain a request from a user, through the visualization tool, that specifies a genomic region represented by the nucleic acid sequencing dataset; and responsive to obtaining the request,

parse the request by: (i) obtaining a plurality of sequencing reads within the genomic region from the nucleic acid sequencing dataset, (ii) running a scan window against the plurality of sequencing reads thereby creating a plurality of windows, each respective window of the plurality of windows corresponding to a different region of the genomic region and including an identity of each identifier of each sequencing read in the different region of the genomic region in the nucleic acid sequencing dataset, and (iii) displaying a two dimensional heat map that represents respective window pairs in the plurality of windows, wherein each respective window pair is displayed in the two dimensional heat map as a color selected from a color scheme based upon the number of identifiers in common in the respective window pair, wherein the plurality of sequencing reads is characterized by a mean fragment length and wherein the number of identifiers in common in a respective window pair is down-weighted to remove bar codes that are expected to be within a multiple of the mean fragment length of each other based on the reference genome sequence corresponding to the sample.

INCORPORATION BY REFERENCE

[0027] All publications, patents, and patent applications mentioned in this specification are herein incorporated by reference in their entireties to the same extent as if each individual publication, patent, or patent application was specifically and individually indicated to be incorporated by reference.

CLAIMS

1. A system for providing structural variation or phasing information over a network connection to a remote client computer, the system comprising one or more microprocessors, a persistent memory and a non-persistent memory, the persistent memory and the non-persistent memory collectively storing one or more nucleic acid sequence datasets, wherein

each respective nucleic acid sequencing dataset in the one or more nucleic acid sequence datasets corresponds to at least one target nucleic acid in a respective sample in a plurality of samples, wherein

the respective sample is associated with a genome of at least one species, the respective nucleic acid sequencing dataset comprises (i) a header, (ii) a synopsis, and (iii) a data section,

the data section comprises a plurality of sequencing reads,

each respective sequencing read in the plurality of sequencing reads comprises a first portion that corresponds to a subset of at least one target nucleic acid in the respective sample and a second portion that encodes a respective identifier for the respective sequencing read in a plurality of identifiers,

each respective identifier is independent of the sequence of the at least one target nucleic acid, and

the plurality of sequencing reads collectively includes the plurality of identifiers, and wherein

the persistent memory and the non-persistent memory further collectively store one or more programs that use the one or more microprocessors to:

provide a visualization tool for installation on the remote client computer;

obtain a request, sent from the remote client computer from a user, over a network connection, for structural variation or phasing information using a first dataset in the one or more datasets, wherein the request is formatted in accordance with an expression syntax; and

responsive to obtaining the request, automatically parse the request by:

- (i) loading the header and the synopsis of the first dataset into the non-persistent memory if not already loaded into the non-persistent memory while retaining the data section in persistent memory,
- (ii) comparing the request to the synopsis of the first dataset thereby identifying one or more portions of the data section of the first dataset,
- (iii) loading the one or more identified portions of the data section into non-persistent memory, wherein the loading loads less than the entirety of the data section,
- (iv) formatting structural variation or phasing information for display on the client computer using the first dataset, and
- (v) transmitting the formatted structural variation or phasing information over the network connection to the remote client computer for display on the remote client computer.

2. A system for providing structural variation or phasing information, the system comprising one or more microprocessors, a persistent memory and a non-persistent memory, the persistent memory and the non-persistent memory collectively storing one or more nucleic acid sequence datasets, wherein

each respective nucleic acid sequencing dataset in the one or more nucleic acid sequence datasets corresponds to at least one target nucleic acid in a respective sample in a plurality of samples, wherein

the respective sample is associated with a genome of at least one species, the respective nucleic acid sequencing dataset comprises (i) a header, (ii) a synopsis, and (iii) a data section,

the data section comprises a plurality of sequencing reads,

each respective sequencing read in the plurality of sequencing reads comprises a first portion that corresponds to a subset of at least one target nucleic acid in the respective sample and a second portion that encodes a respective identifier for the respective sequencing read in a plurality of identifiers,

each respective identifier is independent of the sequence of the at least one target nucleic acid, and

the plurality of sequencing reads collectively includes the plurality of identifiers, and wherein

the persistent memory and the non-persistent memory further collectively store one or more programs that use the one or more microprocessors to:

provide a visualization tool;

obtain a request from a user, through the visualization tool, for structural variation or phasing information using a first dataset in the one or more datasets, wherein the request is formatted in accordance with an expression syntax,

responsive to obtaining the request, automatically parse the request by:

(i) loading the header and the synopsis of the first dataset into the non-persistent memory if not already loaded into the non-persistent memory while retaining the data section in persistent memory,

(ii) comparing the request to the synopsis of the first dataset thereby identifying one or more portions of the data section of the first dataset,

(iii) loading the one or more identified portions of the data section into non-persistent memory, wherein the loading loads less than the entirety of the data section,

(iv) formatting structural variation or phasing information for display in the visualization tool using the first dataset, and

(v) displaying the formatted structural variation or phasing information in the visualization tool.

3. A system for obtaining structural variation or phasing information over a network connection from a remote computer, wherein the system comprises one or more microprocessors, and a memory that stores one or more programs, wherein the one or more programs use the one or more microprocessors to execute a method comprising:

(A) invoking a visualization tool;

(B) obtaining, through the visualization tool from a user, a request for structural variation or phasing information in a first nucleic acid sequencing dataset from among one or more nucleic acid sequencing datasets stored on the remote computer, wherein each respective nucleic acid sequencing dataset in the one or more nucleic acid sequence datasets corresponds to at least one target nucleic acid in a respective sample in a plurality of samples, wherein

the request is formatted in accordance with an expression syntax,

the respective sample is associated with a genome of at least one species, the respective nucleic acid sequencing dataset comprises (i) a header, (ii) a synopsis, and (iii) a data section,

the data section comprises a plurality of sequencing reads,

each respective sequencing read in the plurality of sequencing reads comprises a first portion that corresponds to a subset of at least one target nucleic acid in the respective

sample and a second portion that encodes a respective identifier for the respective sequencing read in a plurality of identifiers,

each respective identifier is independent of the sequence of the at least one target nucleic acid, and

the plurality of sequencing reads collectively includes the plurality of identifiers;

(C) sending the request to the remote computer over the network connection, wherein the remote computer has persistent memory and non-persistent memory, thereby causing the remote computer to execute a method comprising:

(i) loading the header and the synopsis of the first dataset into the non-persistent memory if not already loaded into the non-persistent memory of the remote computer while retaining the data section in persistent memory,

(ii) comparing the request to the synopsis of the first dataset thereby identifying one or more portions of the data section of the first dataset,

(iii) loading the one or more identified portions of the data section into non-persistent memory, wherein the loading loads less than the entirety of the data section, and

(iv) formatting structural variation or phasing information; and

(D) receiving the formatted structural variation or phasing information over the network connection from the remote computer for display in the visualization tool.

4. A system for providing structural variation or phasing information, the system comprising one or more microprocessors, and a memory, wherein the system has access to one or more nucleic acid sequence datasets, wherein

each respective nucleic acid sequencing dataset in the one or more nucleic acid sequence datasets corresponds to at least one target nucleic acid in a respective sample in a plurality of samples, wherein

the respective sample is associated with a genome of at least one species,

the respective nucleic acid sequencing dataset comprises (i) a header, (ii) a synopsis, and (iii) a data section,

the data section comprises a plurality of sequencing reads,

each respective sequencing read in the plurality of sequencing reads comprises a first portion that corresponds to a subset of at least one target nucleic acid in the respective sample and a second portion that encodes a respective identifier for the respective sequencing read in a plurality of identifiers,

each respective identifier is independent of the sequence of the at least one target nucleic acid, and

the plurality of sequencing reads collectively includes the plurality of identifiers, and wherein

the memory further collectively stores one or more programs that use the one or more microprocessors to:

provide a visualization tool;

obtain a request from a user, through the visualization tool, for structural variation or phasing information using a first dataset in the one or more datasets, wherein the request is formatted in accordance with an expression syntax,

responsive to the obtained request, automatically parse the request by:

(i) comparing the request to the synopsis of the first dataset thereby identifying one or more portions of the data section of the first dataset,

(ii) formatting structural variation or phasing information for display in the visualization tool using the identified one or more portions of the data section of the first dataset, wherein the one or more portions of the data section of the first dataset is less than the entirety of the data section of the first dataset, and

(iii) displaying the formatted structural variation or phasing information in the visualization tool.

5. The system of any one of claims 1-4, wherein the header delineates a plurality of components in the respective nucleic acid sequencing dataset.

6. The system of claim 5, wherein the plurality of components comprises two or more components selected from the group consisting of a summary, an index to variant call data, a phase block track, a refseq index track, a gene track, an exon track, an index to read data, a structural variant dataset track, an index to a target dataset, and an index to a fragment dataset.

7. The system of claim 6, wherein the plurality of components comprises the summary and wherein the summary comprises two or more items in the group consisting of:

- a percentage of known SNPs phased in the respective nucleic acid sequencing dataset, a longest phase block in the respective nucleic acid sequencing dataset,
- a number of unique barcodes used in the respective nucleic acid sequencing dataset, an average fragment length in the respective nucleic acid sequencing dataset,
- a mean of the average fragment length in the respective nucleic acid sequencing dataset, a percentage of fragments greater than a lower threshold in the respective nucleic acid sequencing dataset,
- a fragment length histogram in the respective nucleic acid sequencing dataset, an N50 phase block size in the respective nucleic acid sequencing dataset,
- a phase block histogram in the respective nucleic acid sequencing dataset,
- a number of sequence reads represented by respective the nucleic acid sequencing dataset,
- a median insert size in the respective nucleic acid sequencing dataset, a median depth in the respective nucleic acid sequencing dataset,
- a percent of the target genome with zero coverage in the respective nucleic acid sequencing dataset,
- a mapped reads percentage for the respective nucleic acid sequencing dataset,
- a PCR duplication percentage for the respective nucleic acid sequencing dataset,

a coverage histogram for the in the respective nucleic acid sequencing dataset,
an identity of a test nucleic acid that forms the basis for the respective nucleic acid sequencing dataset,
a genome source for the respective nucleic acid sequencing dataset,
a sex of an organism that originated the at least one test nucleic acid of the respective nucleic acid sequencing dataset,
a sex of the organism that originate the respective sample of the in the respective nucleic acid sequencing dataset,
a dataset file format version of the in the respective nucleic acid sequencing dataset, and a
pointer to a plurality of structural variant calls made for the respective nucleic acid sequencing dataset.

8. The system of claim 6, wherein the plurality of components comprises the index to variant call data that provides a correspondence between respective ranges of the genome of the species to offsets in the data section where variant call data for the respective ranges is found.

9. The system of claim 6, wherein the plurality of components comprises the phase block track and wherein the phase block track comprises (i) a dictionary and (ii) a track data section comprising phase information for one or more chromosomes in the genome of the at least one species.

10. The system of claim 9, wherein the dictionary comprises a plurality of names, and for each respective name in the plurality of names, an offset into the track data where records for the corresponding name are found.

11. The system of claim 9, wherein the track data section comprises a plurality of records and wherein each record in the plurality of records represents a phase block in the target nucleic acid.

12. The system of claim 11, wherein the tract data section is in the JSON file format.
13. The system of claim 11, wherein each respective record in the plurality of records specifies:
- (i) a chromosome number corresponding to the respective record,
 - (ii) a position where the phase block starts on the chromosome, (iii) a position where the phase block ends,
 - (iv) a unique name for the record, and
 - (v) phasing information about the phase block.
14. The system of claim 11, wherein each respective record in the plurality of records is represented by a node in a plurality of nodes in a respective interval tree in a plurality of interval trees, wherein each interval tree in the plurality of interval trees represents a chromosome in a plurality of chromosomes for the at least one species.
15. The system of claim 14, wherein
- a node in the plurality of nodes of a first interval tree in the plurality of interval trees stores a midpoint of the node,
 - the midpoint of the node is a position of the midpoint, on the corresponding chromosome, of the phase block corresponding to the node,
 - each respective node in the plurality of nodes of the first interval tree has a link to a left child node, which corresponds to the phase block immediately to the left of the phase block represented by the respective node in the genome of the at least one species,
 - each respective node in the plurality of nodes of the first interval tree has a link to a right child node, which corresponds to the phase block immediately to the right of the phase block represented by the respective node in the genome of the at least one species,
 - each respective node in the plurality of nodes of the first interval tree has a sorted set of nodes that represent phase blocks that overlap the midpoint of the respective node sorted by left hand position of such phase block, and

each respective node in the plurality of nodes of the first interval tree has a sorted set of nodes that represent phase blocks that overlap the midpoint of the respective node sorted by right hand position of such phase blocks.

16. The system of claim 15, wherein each respective node in the plurality of nodes of the first interval tree further includes a name, which is an offset in the track data section to the record in the plurality of records that contains phase information for the phase block corresponding to the respective node.

17. A system for providing structural variation or phasing information over a network connection to a remote client computer, the system comprising one or more microprocessors, a persistent memory and a non-persistent memory, the persistent memory and the non-persistent memory collectively storing one or more nucleic acid sequence datasets, wherein

each respective nucleic acid sequencing dataset in the one or more nucleic acid sequence datasets corresponds to at least one target nucleic acid in a respective sample in a plurality of samples, wherein

the respective sample is associated with a genome of at least one species, the respective nucleic acid sequencing dataset comprises (i) a header, (ii) a synopsis, and (iii) a data section, wherein the header delineates a plurality of components in the respective nucleic acid sequencing dataset and wherein the header further comprises the version of the dataset structure used by the nucleic acid sequencing dataset,

the data section comprises a plurality of sequencing reads,

each respective sequencing read in the plurality of sequencing reads comprises a first portion that corresponds to a subset of at least one target nucleic acid in the respective sample and a second portion that encodes a respective identifier for the respective sequencing read in a plurality of identifiers,

each respective identifier is independent of the sequence of the at least one target nucleic acid, and

the plurality of sequencing reads collectively includes the plurality of identifiers, and wherein

the persistent memory and the non-persistent memory further collectively store one or more programs that use the one or more microprocessors to:

provide a visualization tool for installation on the remote client computer;

obtain a request, sent from the remote client computer from a user, over a network connection, for structural variation or phasing information using a first dataset in the one or more datasets, and

responsive to obtaining the request, automatically parse the request by:

(i) loading the header and the synopsis of the first dataset into the non-persistent memory if not already loaded into the non-persistent memory while retaining the data section in persistent memory,

(ii) comparing the request to the synopsis of the first dataset thereby identifying one or more portions of the data section of the first dataset,

(iii) loading the one or more identified portions of the data section into non-persistent memory, wherein the loading loads less than the entirety of the data section,

(iv) formatting structural variation or phasing information for display on the client computer using the first dataset, and

(v) transmitting the formatted structural variation or phasing information over the network connection to the remote client computer for display on the remote client computer.

18. The system of claim 6, wherein the plurality of components comprises the refseq index, wherein the refseq index comprises an index of a plurality of molecular variation identifiers that are called in the sample.
19. The system of claim 18, wherein each respective molecular variation identifier in the plurality of molecular variation identifiers is dbSNP identifier.
20. The system of claim 6, wherein the plurality of components comprises the gene track, wherein the gene track comprises a plurality of genes and, for each respective gene in the plurality of genes, a number of single nucleotide polymorphisms in the respective gene.
21. The system of claim 6, wherein the plurality of components comprises the gene track and wherein the gene track comprises (i) a gene track dictionary and (ii) a gene track data section.
22. The system of claim 21, wherein the gene track dictionary comprises a plurality of names, and for each respective name in the plurality of names, an offset into the gene track data section where records for the corresponding name are found.
23. The system of claim 21, wherein the gene track data section comprises a plurality of records and wherein each record in the plurality of records represents a gene in the target nucleic acid.
24. The system of claim 23, wherein the gene tract data section is in the JSON file format.
25. The system of claim 23, wherein each respective record in the plurality of records specifies:
- (i) a chromosome number corresponding to the respective record,
 - (ii) a position on the chromosome where the gene represented by the respective record begins,

ends,
 (iii) a position on the chromosome where the gene represented by the respective record
 (iv) a unique name for the respective record, and
 (v) genetic information about the gene.

26. The system of claim 23 wherein the genetic information about the gene comprises an alternative name for the gene, a count of single nucleotide polymorphisms on the gene, or direction of the gene on the respective chromosome.

27. The system of claim 23, wherein each respective gene record in the plurality of gene records is represented by a node in a plurality of nodes in a respective interval tree in a plurality of interval trees, wherein each interval tree in the plurality of interval trees represents a chromosome in a plurality of chromosomes for the at least one species.

28. The system of claim 27, wherein

 a node in the plurality of nodes of a first interval tree in the plurality of interval trees stores a midpoint of the node,

 the midpoint of the node is a position of the midpoint, on the corresponding chromosome, of the gene corresponding to the node,

 each respective node in the plurality of nodes of the first interval tree has a link to a left child node, which corresponds to the gene immediately to the left of the gene represented by the respective node in the genome of the at least one species,

 each respective node in the plurality of nodes of the first interval tree has a link to a right child node, which corresponds to the gene immediately to the right of the gene represented by the respective node in the genome of the at least one species,

 each respective node in the plurality of nodes of the first interval tree has a sorted set of nodes that represent genes that overlap the midpoint of the respective node sorted by left hand position of the respective gene, and

each respective node in the plurality of nodes of the first interval tree has a sorted set of nodes that represent genes that overlap the midpoint of the respective node sorted by right hand position of such genes.

29. The system of claim 28, wherein each respective node in the plurality of nodes of the first interval tree further includes a name, which is an offset in the track data section to the gene record in the plurality of gene records that contains genetic information for the gene corresponding to the respective node.

30. A system for providing structural variation or phasing information, the system comprising one or more microprocessors, a persistent memory and a non-persistent memory, the persistent memory and the non-persistent memory collectively storing one or more nucleic acid sequence datasets, wherein

each respective nucleic acid sequencing dataset in the one or more nucleic acid sequence datasets corresponds to at least one target nucleic acid in a respective sample in a plurality of samples, wherein

the respective sample is associated with a genome of at least one species, the respective nucleic acid sequencing dataset comprises (i) a header, (ii) a synopsis, and (iii) a data section wherein the header delineates a plurality of components in the respective nucleic acid sequencing dataset wherein the plurality of components comprises an index to read data wherein the index to read data comprises a lookup table between a respective identifier in the plurality of identifiers and a shortened version of the respective identifier and wherein the plurality of components further comprises at least one of a summary, an index to variant call data, a phase block track, a refseq index track, a gene track, an exon track, a structural variant dataset track, an index to a target dataset, and an index to a fragment dataset,

the data section comprises a plurality of sequencing reads,

each respective sequencing read in the plurality of sequencing reads comprises a first portion that corresponds to a subset of at least one target nucleic acid in the respective sample and a second portion that encodes a respective identifier for the respective sequencing read in a plurality of identifiers,

each respective identifier is independent of the sequence of the at least one target nucleic acid, and

the plurality of sequencing reads collectively includes the plurality of identifiers, and wherein

the persistent memory and the non-persistent memory further collectively store one or more programs that use the one or more microprocessors to:

provide a visualization tool;

obtain a request from a user, through the visualization tool, for structural variation or phasing information using a first dataset in the one or more datasets,

responsive to obtaining the request, automatically parse the request by:

(i) loading the header and the synopsis of the first dataset into the non-persistent memory if not already loaded into the non-persistent memory while retaining the data section in persistent memory,

(ii) comparing the request to the synopsis of the first dataset thereby identifying one or more portions of the data section of the first dataset,

(iii) loading the one or more identified portions of the data section into non-persistent memory, wherein the loading loads less than the entirety of the data section,

(iv) formatting structural variation or phasing information for display in the visualization tool using the first dataset, and

(v) displaying the formatted structural variation or phasing information in the visualization tool.

31. The system of claim 30, wherein

the data section is stored as a blocked index of records organized into a plurality of chunks, each respective chunk in the plurality of chunks comprises a subset of the plurality of sequencing reads, and

the index to read data further comprises a per chromosome array of chromosome-offset to file-offset associations between the plurality of chunks and absolute positions in the genome of the at least one species that thereby assign each chunk in the plurality of chunks a different absolute position in genome of a species in the at least one species.

32. The system of claim 31, wherein

a respective chunk in the plurality of chunks is an array of structures, each respective structure in the array representing a corresponding sequencing read in the plurality of sequence reads in the respective chunk, and wherein each structure includes:

a first bit reserved to indicate read quality of the corresponding sequencing read, a second bit to represent origin from a first parental haplotype for the corresponding sequencing read,

a third bit to represent origin from a second parental haplotype for the corresponding sequencing read,

a first plurality of bits to represent the shortened version of an identifier in the plurality of identifiers for the corresponding sequencing read,

a second plurality of bits to represent a length of the corresponding sequencing read, and

a third plurality of bits to represent a start position of the corresponding sequencing read relative to the absolute position in genome of the species assigned to the respective chunk.

33. The system of claim 32, wherein the species is human and the respective chunk represents about 1 million base pairs or less.

34. The system of claim 32, wherein the first plurality of bit is 20 bits, the species is human, the respective chunk represents about 1 million base pairs or less, and each structure in the array of structures is a predetermined bit size.

35. The system of claim 32, wherein the identifier, stored in the index to read data, that corresponds to the shortened version of the identifier requires 24 bits.

36. The system of claim 32, wherein the identifier, stored in the index to read data, that corresponds to the shortened version of the identifier requires 30 bits, 32 bits, 34 bits, or 36 bits.

37. The system of claim 6, wherein
the plurality of components comprises the structural variant dataset track, and
the structural variant dataset track comprises (i) a dictionary and (ii) a track data section comprising structural variant call information identified in the plurality of sequencing reads.

38. The system of claim 37, wherein the dictionary comprises a plurality of names, and for each respective name in the plurality of names, an offset into the track data where records for the corresponding name are found.

39. The system of claim 38, wherein a name in the plurality of names corresponds to a chromosome.

40. The system of claim 38, wherein

the track data section comprises a plurality of structural variant records, and
each structural variant record in the plurality of structural variant records represents a structural variant call made in the at least one target nucleic acid in the sample.

41. The system of claim 40, wherein the tract data section is in the JSON file format.

42. The system of claim 40, wherein a respective structural variant record in the plurality of structural variant records specifies a start point on a first chromosome and an end point on a second chromosome for the structural variant represented by the respective structural variant record and at least one on of:

(i) a name of the structural variant represented by the respective structural variant record,
and

(ii) an expression of the quality in the confidence of the identification of the structural variant represented by the respective structural variant record.

43. The system of claim 40, wherein

each respective structural variant record in the plurality of structural variant records is represented by a node in a plurality of nodes in a respective interval tree in a plurality of interval trees, and

each interval tree in the plurality of interval trees represents a chromosome in a plurality of chromosomes for the species.

44. The system of claim 43, wherein

a node in the plurality of nodes of a first interval tree in the plurality of interval trees stores a midpoint of the node,

the midpoint of the node is a position of the midpoint, on the corresponding chromosome, of the structural variant corresponding to the node,

each respective node in the plurality of nodes of the first interval tree has a link to a left child node, which corresponds to the structural variant immediately to the left of the structural variant represented by the respective node in the genome of the species,

each respective node in the plurality of nodes of the first interval tree has a link to a right child node, which corresponds to the structural variant immediately to the right of the structural variant represented by the respective node in the genome of the species,

each respective node in the plurality of nodes of the first interval tree has a sorted set of nodes that represent structural variant that overlap the midpoint of the respective node sorted by left hand position of such structural variant, and

each respective node in the plurality of nodes of the first interval tree has a sorted set of nodes that represent structural variants that overlap the midpoint of the respective node sorted by right hand position of such structural variants.

45. The system of claim 6, wherein

the plurality of components comprises the index to the target dataset,

the target dataset comprises the regions of the at least one target nucleic acid in the sample that were selected for sequencing in the respective nucleic acid sequencing dataset,

the target dataset is indexed by a target dataset index stored in the synopsis, and the target dataset is stored in the data section.

46. The system of claim 45, further comprising instructions, executed by the one or more physical processors to use the target dataset index to calculate which portions of the target dataset must be read from the target dataset to fulfill the request.

47. The system of claim 45, wherein

the target dataset index is split up by chromosome and, for each respective chromosome, and

the target dataset index stores a corresponding array associating ranges on the respective chromosome with the offset at which specific data for that range may be found in the target dataset.

48. The system of claim 6, wherein

the plurality of components comprises the index to the fragment dataset,

the fragment dataset comprises a length, chromosomal position, identifier, and phase of each fragment of the at least one target nucleic acid in the sample,

the fragment dataset is indexed by a fragment dataset index stored in the synopsis, and the fragment dataset is stored in the data section.

49. The system of claim 48, further comprising instructions, executed by the one or more physical processors to use the fragment dataset index to calculate which portion of the fragment dataset must be read from the fragment dataset to fulfill the request.

50. The system of claim 49, wherein the fragment dataset index is split up by chromosome and, for each respective chromosome, the fragment dataset index stores a corresponding array associating ranges on the respective chromosome with the offset at which specific data for that range may be found in the fragment dataset.

51. A system for obtaining phasing information over a network connection from a remote computer, wherein the system comprises one or more microprocessors, and a memory that stores one or more programs, wherein the one or more programs use the one or more microprocessors to execute a method comprising:

(A) invoking a visualization tool;

(B) obtaining, through the visualization tool from a user, a request for phasing information in a first nucleic acid sequencing dataset from among one or more nucleic acid sequencing datasets stored on the remote computer, wherein each respective nucleic acid sequencing dataset in the one or more nucleic acid sequence datasets corresponds to at least one target nucleic acid in a respective sample in a plurality of samples, wherein

the respective sample is associated with a genome of at least one species, the respective nucleic acid sequencing dataset comprises (i) a header, (ii) a synopsis, and (iii) a data section,

the data section comprises a plurality of sequencing reads, each respective sequencing read in the plurality of sequencing reads comprises a first portion that corresponds to a subset of at least one target nucleic acid in the respective sample and a second portion that encodes a respective identifier for the respective sequencing read in a plurality of identifiers,

each respective identifier is independent of the sequence of the at least one target nucleic acid, and

the plurality of sequencing reads collectively includes the plurality of identifiers;

(C) sending the request to the remote computer over the network connection, wherein the remote computer has persistent memory and non-persistent memory, thereby causing the remote computer to execute a method comprising:

(i) loading the header and the synopsis of the first dataset into the non-persistent memory if not already loaded into the non-persistent memory of the remote computer while retaining the data section in persistent memory,

(ii) comparing the request for phasing information to the synopsis of the first dataset thereby identifying one or more portions of the data section of the first dataset,

(iii) loading the one or more identified portions of the data section into non-persistent memory, wherein the loading loads less than the entirety of the data section, and

(iv) formatting phasing information; and

(D) receiving the formatted phasing information over the network connection from the remote computer for display in the visualization tool, wherein the request is for phasing information in a region of the genome and the formatted phasing information includes a graphic representation comprising:

a first haplotype track corresponding to a first parental haplotype of a first species in the at least one species in the region of the genome for the first dataset,

a second haplotype track, corresponding to a second parental haplotype of the first species in the region of the genome for the first dataset, and

an indeterminate track corresponding to regions of the at least one nucleic acid sample that have not been assigned a parental haplotype in the region of the genome for the first dataset.

52. The system of claim 51, wherein each phase block in the region in the first dataset is encompassed by a rectangular box.

53. The system of claim 52, wherein the region in the first dataset comprises two or more phase blocks, five or more phase blocks, or ten or more phase blocks.

54. The system of claim 51, wherein the first haplotype track, the second haplotype track and the indeterminate track includes a plurality of vertical bars, wherein each vertical bar represents a single nucleotide polymorphism, insertion, or deletion, in the at least one target nucleic acid.

55. The system of claim 54, wherein each vertical bar in the plurality of vertical bars is color coded to indicate one of a reference genotype and an alternative genotype.

56. The system of claim 51, wherein the graphic representation further comprises a chromosome map and a location on the chromosome map that is encompassed by the region of the genome.
57. The system of claim 51, wherein the graphic representation further comprises a graphic representation of each gene that is in the region of the genome.
58. The system of claim 51, wherein the graphic representation further comprises a graphic representation of each exon that is in the region of the genome.
59. The system of claim 51, wherein the graphic representation further comprises a coverage track for the region of the genome, wherein the coverage track comprises a plurality of vertical bars, and wherein each respective vertical bar in the plurality of vertical bars indicates an average coverage-per-base in the first dataset for a corresponding portion of the genome under the bar.
60. The system of claim 51, wherein the graphic representation further comprises a breakpoints track for one or more structural variants arising in the region of the genome.
61. The system of claim 60, wherein the breakpoints track wherein the one or more structural variants comprise an inter-chromosomal translocation, a gene fusion, an inversion, or a deletion.
62. The system of claim 60, wherein a first structural variant in the one or more structural variants is formatted as a graphic that, when selected by the user, provides a structural variant zoom affordance that, when selected, provides an additional haplotype track zoomed to the breakpoints of the first structural variant.
63. The system of any one of claims 1-4, wherein the one or more nucleic acid sequencing datasets is a single nucleic acid sequencing dataset.

64. The system of any one of claims 1-4, wherein the one or more nucleic acid sequencing datasets is a plurality of nucleic acid sequencing datasets.

65. The system of claim 64, wherein the plurality of nucleic acid sequencing datasets comprises ten nucleic acid sequencing datasets, 100 nucleic acid sequencing datasets, or 1000 nucleic acid sequencing datasets.

66. The system of any one of claims 1-65, wherein the expression syntax is X1:N1-N2, wherein
X1 is an identity of a selected first chromosome or a selected first contig sequence, N1 is
a selected start position within the first chromosome or the selected first contig
sequence, and
N2 is a selected end position within the first chromosome or the selected first contig
sequence.

67. The system of any one of claims 1-65, wherein the expression syntax is
X1:N1-N2, wherein
X1 is an identity within a selected first chromosome or a selected first contig sequence,
N1 is a selected start position within the first chromosome or the selected first contig
sequence, and
N2 is a selected end position within the first chromosome or the selected first contig
sequence.

68. The system of any one of claims 1-65, wherein the expression syntax is X1:N1, wherein
X1 is an identity of a selected first chromosome or a selected first contig sequence, and
N1 is a number of nucleotides, beginning at the origin of the first chromosome or the
selected first contig sequence.

69. The system of any one of claims 1-65, wherein the expression syntax is Y1, Y2, ..., YN,
wherein
each Yi in Y1, Y2, ..., YN is either an alphanumeric identification of a selected gene, a
selection of a chromosomal region, or selection of a region of a contig sequence.

70. The system of claim 69, wherein
a first Yi in Y1, Y2, ..., YN is an identity of a first chromosome or a first contig sequence
having the syntax X1:N1-N2, wherein X1 is an identity of the first chromosome or the first contig
sequence, N1 is a selected start position within the first chromosome or the first contig sequence,
and N2 is a selected end position within the first chromosome or the first contig sequence, and
a second Yi in Y1, Y2, ..., YN is an alphanumeric identification of a selected gene.

71. The system of claim 69, wherein
a first Yi in Y1, Y2, ..., YN is an identity of a first chromosome or a first contig sequence
having the syntax X1:N1-N2, wherein X1 is an identity of the first chromosome or the first contig
sequence, N1 is a selected start position within the first chromosome or the first contig sequence,
and N2 is a selected end position within the first chromosome or the first contig sequence, and
a second Yi in Y1, Y2, ..., YN is an alphanumeric identification of a selected gene.

72. The system of any one of claims 1-4, wherein the request is converted, without human intervention, to genomic coordinates by comparison of the request against one or more lookup tables that match alphanumeric entries of genes to genomic coordinates.

73. The system of any one of claims 1-4, wherein the request comprises one or more gene names, one or more genomic coordinates, or a combination thereof.

74. A system for providing structural variation or phasing information over a network connection to a remote client computer, the system comprising one or more microprocessors, a persistent memory and a non-persistent memory, the persistent memory and the non-persistent memory collectively storing one or more nucleic acid sequence datasets, wherein

each respective nucleic acid sequencing dataset in the one or more nucleic acid sequence datasets corresponds to at least one target nucleic acid in a respective sample in a plurality of samples, wherein the respective sample is associated with a genome of a plurality of species and includes at least a portion of the genome of a first species and a portion of the genome of the second species and, wherein

the respective sample is associated with a genome of at least one species, the respective nucleic acid sequencing dataset comprises (i) a header, (ii) a synopsis, and (iii) a data section,

the data section comprises a plurality of sequencing reads,

each respective sequencing read in the plurality of sequencing reads comprises a first portion that corresponds to a subset of at least one target nucleic acid in the respective

sample and a second portion that encodes a respective identifier for the respective sequencing read in a plurality of identifiers,

each respective identifier is independent of the sequence of the at least one target nucleic acid, and

the plurality of sequencing reads collectively includes the plurality of identifiers, and wherein

the persistent memory and the non-persistent memory further collectively store one or more programs that use the one or more microprocessors to:

provide a visualization tool for installation on the remote client computer;

obtain a request, sent from the remote client computer from a user, over a network connection, for structural variation or phasing information using a first dataset in the one or more datasets; and

responsive to obtaining the request, automatically parse the request by:

(i) loading the header and the synopsis of the first dataset into the non-persistent memory if not already loaded into the non-persistent memory while retaining the data section in persistent memory,

(ii) comparing the request to the synopsis of the first dataset thereby identifying one or more portions of the data section of the first dataset,

(iii) loading the one or more identified portions of the data section into non-persistent memory, wherein the loading loads less than the entirety of the data section,

(iv) formatting structural variation or phasing information for display on the client computer using the first dataset, and

(v) transmitting the formatted structural variation or phasing information over the network connection to the remote client computer for display on the remote client computer.

75. The system of claim 74, wherein the portion of the genome of the first species is integrated into the portion of the genome of the second species.

76. The system of claim 74, wherein the first species is a retrovirus.

77. The system of claim 74, wherein the portion of the genome of the first species is not integrated into the portion of the genome of the second species.

78. A system for processing program output over a network connection using a local computer, wherein the local computer comprises one or more microprocessors, and a memory that stores one or more programs, wherein the one or more programs use the one or more microprocessors to execute a method in accordance with a first operating system running on the local computer, the method comprising:

(A) invoking a first instance of a first program;

(B) obtaining, through the first instance of the first program from a user, a login and a password to a user account on a remote computer;

(C) logging the user into the user account on the remote computer automatically, using the login and the password provided by the first instance of the first program, across a network connection between the local computer and the remote computer;

(D) responsive to successful login on the remote computer, automatically sending, without human intervention, a second instance of the first program configured to auto-install on the remote computer upon transmission to the remote computer;

(E) receiving from the remote computer a request to open a panel within the first instance of the first program, wherein the panel is originated by the second instance of the first program running on the remote computer, and wherein the panel solicits input from the user for controlling the second instance of the first program;

(F) responsive to receiving input from the user for controlling the second instance of the first program in the panel on the local computer, sending the input to the second instance of the first program on the remote computer; and

(G) receiving, from the remote computer across the network connection, output from the second instance of the first program responsive to the input and displaying the output at the local computer.

79. The system of claim 78, wherein remote computer is a server computer and the local computer is a desktop computer or laptop computer.

80. The system of claim 78, wherein the first operating system is other than the second operating system.

81. The system of claim 78, wherein the receiving (E) further solicits an encryption key from the user and wherein communication between the first instance on the client computer and the second instance on the remote computer is encrypted.

82. The system of claim 78, wherein the remote computer comprises a persistent memory and a non-persistent memory, the persistent memory and the non- persistent memory collectively storing one or more nucleic acid sequence datasets, wherein

each respective nucleic acid sequencing dataset in the one or more nucleic acid sequence datasets corresponds to at least one target nucleic acid in a respective sample in a plurality of samples, wherein

the respective sample is associated with a genome of a species,

the respective nucleic acid sequencing dataset comprises (i) a header, (ii) a synopsis, and (iii) a data section,

the data section comprises a plurality of sequencing reads,

each respective sequencing read in the plurality of sequencing reads comprises a first portion that corresponds to a subset of at least one target nucleic acid in the respective sample and a second portion that encodes a respective identifier for the respective sequencing read in a plurality of identifiers,

each respective identifier is independent of the sequence of the at least one target nucleic acid, and

the plurality of sequencing reads collectively includes the plurality of identifiers, and wherein

the input received from the user for controlling the second instance of the first program is a request for structural variation or phasing information using a first dataset in the one or more datasets, and

responsive to receiving the request, the second instance of the first program filters the request by:

(i) loading the header and the synopsis of the first dataset into the non-persistent memory if not already loaded into the non-persistent memory while retaining the data section in persistent memory,

(ii) comparing the request to the synopsis of the first dataset thereby identifying one or more portions of the data section of the first dataset,

(iii) loading the one or more identified portions of the data section into non-persistent memory,

(iv) formatting structural variation or phasing information for display in the panel in the first instance of the first program using the first dataset, and

(v) transmitting the formatted structural variation or phasing information as the output from the second instance of the first program over the network connection to the local computer for display in the first panel.

83. A system for viewing nucleic acid sequencing data, the system comprising one or more microprocessors and a memory, wherein the memory stores one or more programs that use the one or more microprocessors to:

obtain a nucleic acid sequencing dataset corresponding to at least one target nucleic acid in a sample, the nucleic acid sequencing dataset comprising a plurality of sequencing reads from the sample, wherein

each respective sequencing read in the plurality of sequencing reads comprises a first portion that corresponds to a subset of at least one target nucleic acid in the sample and a second portion that encodes a respective identifier for the respective sequencing read in a plurality of identifiers,

each respective identifier is independent of the sequence of the at least one target nucleic acid, and

the plurality of sequencing reads collectively includes the plurality of identifiers,

display a visualization tool;

obtain a request from a user, through the visualization tool, that specifies a genomic region represented by the nucleic acid sequencing dataset; and

responsive to the obtained request, parse the request by:

(i) obtaining a plurality of sequencing reads within the genomic region from the nucleic acid sequencing dataset,

(ii) running a scan window against the plurality of sequencing reads thereby creating a plurality of windows, each respective window of the plurality of windows corresponding to a different region of the genomic region and including an identity of each identifier of each sequencing read in the different region of the genomic region in the nucleic acid sequencing dataset, and

(iii) displaying a two dimensional heat map that represents each possible window pair in the plurality of windows, wherein each respective window pair is displayed in the two dimensional heat map as a color selected from a color scheme based upon the number of

identifiers in common in the respective window pair, wherein the visualization tool provides a zoom affordance to zoom the heat map or a pan affordance to pan the heat map.

84. The system of claim 83, wherein the visualization tool provides a hide expected overlap affordance, wherein

when a user engages the hide expected overlap affordance, identifiers of sequencing reads that are expected to be adjacent to each other in accordance with a reference genome do not contribute to a count of a number of identifiers in common in the respective window pair, and

when a user does not engage the hide expected overlap affordance, identifiers of sequencing reads that are expected to be adjacent to each other in accordance with a reference genome contribute to the count of the number of identifiers in common in the respective window pair.

85. The system of claim 83, wherein the visualization tool provides a quality affordance, wherein

when a user engages the quality affordance, identifiers of respective sequencing reads that do not satisfy a quality threshold quantifying a quality of assignment of the respective sequence reads to locations in a reference genome do not contribute to a count of the number of identifiers in common in the respective window pair,

when a user does not engage the quality affordance, identifiers of respective sequencing reads that do not satisfy a quality threshold quantifying a quality of assignment of the respective sequence reads to locations in a reference genome contribute to the count of the number of identifiers in common in the respective window pair.

86. A method for providing structural variation or phasing information over a network connection to a remote client computer, the method comprising:

at a system comprising one or more microprocessors, a persistent memory and a non-persistent memory, the persistent memory and the non-persistent memory collectively storing one or more nucleic acid sequence datasets, wherein

each respective nucleic acid sequencing dataset in the one or more nucleic acid sequence datasets corresponds to at least one target nucleic acid in a respective sample in a plurality of samples, wherein

the respective sample is associated with a genome of at least one species, the respective nucleic acid sequencing dataset comprises (i) a header, (ii) a synopsis, and (iii) a data section,

the data section comprises a plurality of sequencing reads,

each respective sequencing read in the plurality of sequencing reads comprises a first portion that corresponds to a subset of at least one target nucleic acid in the respective sample and a second portion that encodes a respective identifier for the respective sequencing read in a plurality of identifiers,

each respective identifier is independent of the sequence of the at least one target nucleic acid, and

the plurality of sequencing reads collectively includes the plurality of identifiers:

(A) providing a visualization tool for installation on the remote client computer;

(B) obtaining a request, sent from the remote client computer from a user, over a network connection, for structural variation or phasing information using a first dataset in the one or more datasets, wherein the request is formatted in accordance with an expression syntax; and

(C) responsive to obtaining the request, automatically parsing the request by:

(i) loading the header and the synopsis of the first dataset into the non-persistent memory if not already loaded into the non-persistent memory while retaining the data section in persistent memory,

(ii) comparing the request to the synopsis of the first dataset thereby identifying one or more portions of the data section of the first dataset,

(iii) loading the one or more identified portions of the data section into non-persistent memory, wherein the loading loads less than the entirety of the data section,

(iv) formatting structural variation or phasing information for display on the client computer using the first dataset, and

(v) transmitting the formatted structural variation or phasing information over the network connection to the remote client computer for display on the remote client computer.

87. A method for providing structural variation or phasing information, the method comprising:

at a system comprising one or more microprocessors, a persistent memory and a non-persistent memory, the persistent memory and the non-persistent memory collectively storing one or more nucleic acid sequence datasets, wherein

each respective nucleic acid sequencing dataset in the one or more nucleic acid sequence datasets corresponds to at least one target nucleic acid in a respective sample in a plurality of samples, wherein

the respective sample is associated with a genome of at least one species,

the respective nucleic acid sequencing dataset comprises (i) a header, (ii) a synopsis, and (iii) a data section,

the data section comprises a plurality of sequencing reads,

each respective sequencing read in the plurality of sequencing reads comprises a first portion that corresponds to a subset of at least one target nucleic acid in the respective sample and a second portion that encodes a respective identifier for the respective sequencing read in a plurality of identifiers,

each respective identifier is independent of the sequence of the at least one target nucleic acid, and

the plurality of sequencing reads collectively includes the plurality of identifiers:

(A) providing a visualization tool;

(B) obtaining a request from a user, through the visualization tool, for structural variation or phasing information using a first dataset in the one or more datasets wherein the request is formatted in accordance with an expression syntax,

(C) responsive to obtaining the request, automatically parsing the request by:

(i) loading the header and the synopsis of the first dataset into the non-persistent memory if not already loaded into the non-persistent memory while retaining the data section in persistent memory,

(ii) comparing the request to the synopsis of the first dataset thereby identifying one or more portions of the data section of the first dataset,

(iii) loading the one or more identified portions of the data section into non-persistent memory, wherein the loading loads less than the entirety of the data section,

(iv) formatting structural variation or phasing information for display in the visualization tool using the first dataset, and

(v) displaying the formatted structural variation or phasing information in the visualization tool.

88. A method for obtaining structural variation or phasing information over a network connection from a remote computer, the method comprising:

at a system comprising one or more microprocessors, and a memory that stores one or more programs:

(A) invoking a visualization tool;

(B) obtaining, through the visualization tool from a user, a request for structural variation or phasing information in a first nucleic acid sequencing dataset from among one or more nucleic acid sequencing datasets stored on the remote computer, wherein each respective nucleic acid sequencing dataset in the one or more nucleic acid sequence datasets corresponds to at least one target nucleic acid in a respective sample in a plurality of samples, wherein

the request is formatted in accordance with an expression syntax,

the respective sample is associated with a genome of at least one species, the

respective nucleic acid sequencing dataset comprises (i) a header, (ii) a synopsis, and (iii) a data section,

the data section comprises a plurality of sequencing reads,

each respective sequencing read in the plurality of sequencing reads comprises a first portion that corresponds to a subset of at least one target nucleic acid in the respective sample and a second portion that encodes a respective identifier for the respective sequencing read in a plurality of identifiers,

each respective identifier is independent of the sequence of the at least one target nucleic acid, and

the plurality of sequencing reads collectively includes the plurality of identifiers;

(C) sending the request to the remote computer over the network connection, wherein the remote computer has persistent memory and non-persistent memory, thereby causing the remote computer to execute a method comprising:

(i) loading the header and the synopsis of the first dataset into the non-persistent memory if not already loaded into the non-persistent memory of the remote computer while retaining the data section in persistent memory,

(ii) comparing the request to the synopsis of the first dataset thereby identifying one or more portions of the data section of the first dataset,

(iii) loading the one or more identified portions of the data section into non-persistent memory, wherein the loading loads less than the entirety of the data section, and

(iv) formatting structural variation or phasing information; and

(D) receiving the formatted structural variation or phasing information over the network connection from the remote computer for display in the visualization tool.

89. A method for providing structural variation or phasing information, the method comprising:

at a system comprising one or more microprocessors, and a memory, wherein the system has access to one or more nucleic acid sequence datasets, wherein

each respective nucleic acid sequencing dataset in the one or more nucleic acid sequence datasets corresponds to at least one target nucleic acid in a respective sample in a plurality of samples, wherein

the respective sample is associated with a genome of at least one species,

the respective nucleic acid sequencing dataset comprises (i) a header, (ii) a synopsis, and (iii) a data section,

the data section comprises a plurality of sequencing reads,

each respective sequencing read in the plurality of sequencing reads comprises a first portion that corresponds to a subset of at least one target nucleic acid in the respective sample and a second portion that encodes a respective identifier for the respective sequencing read in a plurality of identifiers,

each respective identifier is independent of the sequence of the at least one target nucleic acid, and

the plurality of sequencing reads collectively includes the plurality of identifiers,

(A) providing a visualization tool;

(B) obtaining a request from a user, through the visualization tool, for structural variation or phasing information using a first dataset in the one or more datasets, wherein the request is formatted in accordance with an expression syntax; and

(C) responsive to obtaining the request, automatically parsing the request by a method comprising:

(i) comparing the request to the synopsis of the first dataset thereby identifying one or more portions of the data section of the first dataset,

(ii) formatting structural variation or phasing information for display in the visualization tool using the identified one or more portions of the data section of the first dataset, wherein the one or more portions of the data section of the first dataset is less than the entirety of the data section of the first dataset, and

(iii) displaying the formatted structural variation or phasing information in the visualization tool.

90. A method for processing program output over a network connection, the method comprising:

at a local computer, wherein the local computer comprises one or more microprocessors, and a memory that stores one or more programs to:

(A) invoke a first instance of a first program;

(B) obtain, through the first instance of the first program from a user, a login and a password to a user account on a remote computer;

(C) log the user into the user account on the remote computer automatically, using the login and the password provided by the first instance of the first program, across a network connection between the local computer and the remote computer;

(D) responsive to successful login on the remote computer, automatically send, without human intervention, a second instance of the first program configured to auto-install on the remote computer upon transmission to the remote computer;

(E) receive from the remote computer a request to open a panel within the first instance of the first program, wherein the panel is originated by the second instance of the first program running on the remote computer, and wherein the panel solicits input from the user for controlling the second instance of the first program;

(F) responsive to receiving input from the user for controlling the second instance of the first program in the panel on the local computer, send the input to the second instance of the first program on the remote computer; and

(G) receive, from the remote computer across the network connection, output from the second instance of the first program responsive to the input and displaying the output at the local computer.

91. A method for viewing nucleic acid sequencing data, the method comprising:

at a system comprising one or more microprocessors and a memory, wherein the memory stores one or more programs that use the one or more microprocessors to:

obtain a nucleic acid sequencing dataset corresponding to at least one target nucleic acid in a sample, the nucleic acid sequencing dataset comprising a plurality of sequencing reads from the sample, wherein

each respective sequencing read in the plurality of sequencing reads comprises a first portion that corresponds to a subset of at least one target nucleic acid in the sample and a second portion that encodes a respective identifier for the respective sequencing read in a plurality of identifiers,

each respective identifier is independent of the sequence of the at least one target nucleic acid, and

the plurality of sequencing reads collectively includes the plurality of identifiers,

display a visualization tool;

obtain a request from a user, through the visualization tool, that specifies a genomic region represented by the nucleic acid sequencing dataset; and

responsive to obtaining the request, parse the request by:

(i) obtaining a plurality of sequencing reads within the genomic region from the nucleic acid sequencing dataset,

(ii) running a scan window against the plurality of sequencing reads thereby creating a plurality of windows, each respective window of the plurality of windows corresponding to a different region of the genomic region and including an identity of each identifier of each sequencing read in the different region of the genomic region in the nucleic acid sequencing dataset, and

(iii) displaying a two dimensional heat map that represents respective window pairs in the plurality of windows, wherein each respective window pair is displayed in the two dimensional heat map as a color selected from a color scheme based upon the number of identifiers in common in the respective window pair, wherein the plurality of sequencing reads is characterized by a mean fragment length and wherein the number of identifiers in common in a respective window pair is down-weighted to remove bar codes that are expected to be within a

multiple of the mean fragment length of each other based on the reference genome sequence corresponding to the sample.

92. The method of claim 91 wherein a window pair in the two dimensional heat map represents a first window and a second window that are more than 100 kilobases apart from each other in a reference genome corresponding to the sample.

93. The method of claim 91 wherein a window pair in the two dimensional heat map represents a first window and a second window that are more than a megabase apart from each other in a reference genome corresponding to the sample.

94. The method of claim 91 wherein a window pair in the two dimensional heat map represents a first window and a second window that are more than 5 megabases apart from each other in a reference genome corresponding to the sample.

95. The method of claim 91, wherein the multiple is a real number between 0 and 1.

96. The method of claim 91, wherein the multiple is a real number of 1 or greater.

97. The method of claim 91, wherein the multiple is a positive integer.