

fragments that are in the same space. Thus ligated DNA fragments are representative of the genomic environment from which the fragments originate. By determining at least part of the sequence of all the different ligated DNA fragments, DNA fragment sequences may subsequently be coupled using the sequence information of the different ligated DNA

5 fragments and a sequence for separate genomic regions of interest may be built.

In one aspect, there is provided a method for building a contig of a genomic region of interest comprising a target nucleotide sequence comprising, fragmenting a crosslinked DNA, ligating the fragmented cross linked DNA, reversing the crosslinking and determining at least part of the sequences of ligated DNA fragments which comprise a target nucleotide sequence, and
10 using the determined sequences to build a contig of the genomic region of interest.

In another aspect, there is provided a method for building a contig of a genomic region of interest comprising a target nucleotide sequence, comprising the steps of:

- a) providing a sample of cross-linked DNA;
- b) fragmenting the crosslinked DNA;
- 15 c) ligating the fragmented crosslinked DNA;
- d) reversing the crosslinking;
- e) optionally fragmenting the DNA of step d), preferably with a restriction enzyme;
- f) optionally, ligating the fragmented DNA of step d) or e) to at least one Adaptor;
- 20 g) optionally and preferably, amplifying the ligated DNA fragments of step d) or e)
comprising the target nucleotide sequence using at least one primer which hybridises
to the target nucleotide sequence, or amplifying the ligated DNA fragments of step f)
using at least one primer which hybridises to the target nucleotide sequence and at
least one primer which hybridises to the at least one adaptor;
- 25 h) determining at least part of the sequence of the (amplified) ligated DNA
fragments of step d), e), f) or g) comprising the target nucleotide sequence preferably
using high throughput sequencing;
- i) building a contig of the genomic region of interest from the determined
sequences.

In yet another aspect, there is provided a method for building a contig of a genomic region of
30 interest comprising a target nucleotide sequence, comprising the steps of:

- a) providing a sample of cross-linked DNA;
- b) fragmenting the crosslinked DNA;
- c) ligating the fragmented crosslinked DNA;
- d) reversing the crosslinking;
- 35 e) optionally fragmenting the DNA of step d), preferably with a restriction enzyme
- f) circularising the DNA of step d) or e);

- g) optionally and preferably, amplifying the circularised DNA comprising the target nucleotide sequence using preferably at least one primer which hybridises to the target nucleotide sequence;
- h) determining at least part of the sequence of the (amplified) ligated DNA fragments comprising the target nucleotide using high throughput sequencing;
- i) building a contig of the genomic region of interest from the determined sequences.

In yet another aspect, there is provided a method for identifying the presence or absence of a genetic mutation, comprising the steps a)-i) as defined herein, wherein contigs are built for a plurality of samples, comprising the further steps of:

- j) aligning the contigs of a plurality of samples;
- k) identifying the presence or absence of a genetic mutation in the genomic regions of interest from the plurality of samples.

In yet another aspect, there is provided a method for identifying the presence or absence of a genetic mutation, comprising the steps a)-i) as defined herein, comprising the further steps of:

- l) aligning the contig to a reference sequence.
- m) identifying the presence or absence of a genetic mutation in the genomic region of interest.

Definitions

In the following description and examples, a number of terms are used. In order to provide a clear and consistent understanding of the specification and claims, including the scope to be given such terms, the following definitions are provided. Unless otherwise defined herein, all technical and scientific terms used have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. The disclosures of all publications, patent applications, patents and other references are incorporated herein in their entirety by reference.

Methods of carrying out the conventional techniques used in methods of the invention will be evident to the skilled worker. The practice of conventional techniques in molecular biology, biochemistry, computational chemistry, cell culture, recombinant DNA, bioinformatics, genomics, sequencing and related fields are well-known to those of skill in the art and are discussed, for example, in the following literature references: Sambrook et al. ., Molecular Cloning. A Laboratory Manual, 2nd Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N. Y., 1989; Ausubel et al., Current Protocols in Molecular Biology, John Wiley & Sons, New York, 1987 and periodic updates; and the series Methods in Enzymology, Academic Press, San Diego.

As used herein, the singular forms "a," "an" and "the" include plural referents unless the context clearly dictates otherwise. For example, a method for isolating "a" DNA molecule, as used above, includes isolating a plurality of molecules (e.g. 10's, 100's, 1000's, 10's of thousands, 100's of thousands, millions, or more molecules).

- 5 A "genomic region of interest" according to the invention is a DNA sequence of an organism of which it is desirable to determine, at least part of, the DNA sequence. For instance, a genomic region which is suspected of comprising an allele associated with a disease may be a genomic region of interest. As used herein, the term "allele(s)" means any of one or more alternative forms of a gene at a particular locus. In a diploid cell of an
- 10 organism, alleles of a given gene are located at a specific location, or locus (loci plural) on a chromosome. One allele is present on each chromosome of the pair of homologous chromosomes. Thus, in a diploid cell, two alleles and thus two separate (different) genomic regions of interest may exist.