

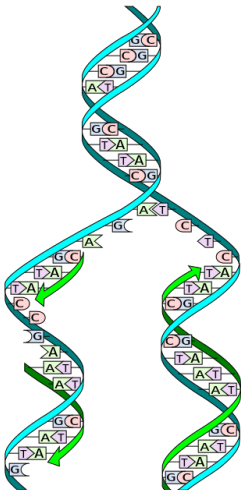
Systematyczne przetwarzanie informacji o rearanżacjach genomowych w R

Piotr Dittwald (MISDoMP/MIMUW)

piotr.dittwald@gmail.com

SER, Warszawa, 22 V 2014

Ludzki genom

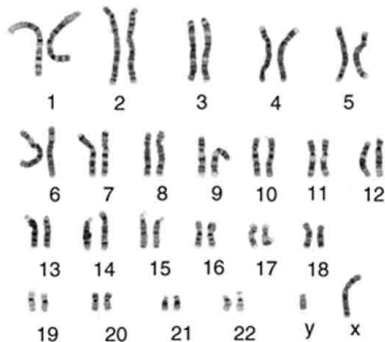


źródło: wikipedia

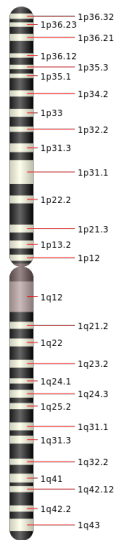
- ▶ cztery możliwe nukleotydy: adenina (A), cytozyna (C), tymina (T), guanina (G)
- ▶ dwie helisy DNA
- ▶ komplementarne pary nukleotydów: A-T, G-C

Karyotyp

- ▶ chromosomy 1-22 (autosomy)
- ▶ chromosomy płciowe (mężczyzna: X-Y; kobieta: X-X)
- ▶ ramiona p i q, centromer, prążki widoczne po zabarwieniu barwnikiem



źródło: http://www.biotechnologyonline.gov.au/popups/img_karyotype.html



źródło: <http://ghr.nlm.nih.gov/chromosome=1>

- ▶ repozytorium pakietów do bioinformatyki
- ▶ na stronie dostępna dokumentacja i tzw. *vignette*

The screenshot shows a web browser window displaying the Bioconductor website. The page is for the 'IRanges' package, version 2.14. The header includes the Bioconductor logo and navigation links: Home, Install, Help, Developers, and About. The main content area is titled 'IRanges' and describes it as 'Infrastructure for manipulating intervals on sequences'. It provides information about the package version (Release (2.14)), its purpose (efficient low-level and highly reusable S4 classes for storing ranges of integers, RLE vectors, and more generally, data that can be organized sequentially), and its author (H. Pages, P. Abouyon and M. Lawrence). The page also includes a 'source' code block for installing the package, a 'citation' code block for citing the package, and a 'Documentation' section with links to PDF files and R scripts. A 'Details' section at the bottom shows the package's version (1.22.6) and its dependencies (BioC 2.3 (R-2.8)).

Bioconductor - IRanges - Mozilla Firefox

Firefox - Inbox - Plotr Di... Cytogen... gene de... Bioc... Table B... Genom... http...rja barplot... R-cran p... R Q

www.bioconductor.org/packages/release/bioc/html/IRanges.html

IRanges

Search:

Bioconductor
OPEN SOURCE SOFTWARE FOR BIOINFORMATICS

Home Install Help Developers About

Home » Bioconductor 2.14 » Software Packages » IRanges

IRanges

Infrastructure for manipulating intervals on sequences

Bioconductor version: Release (2.14)

The package provides efficient low-level and highly reusable S4 classes for storing ranges of integers, RLE vectors (Run-Length Encoding), and, more generally, data that can be organized sequentially (formally defined as Vector objects), as well as views on these Vector objects. Efficient list-like classes are also provided for storing big collections of instances of the basic classes. All classes in the package use consistent naming and share the same rich and consistent "Vector API" as much as possible.

Author: H. Pages, P. Abouyon and M. Lawrence

Maintainer: Bioconductor Package Maintainer <maintainer at bioconductor.org>

To install this package, start R and enter:

```
source("http://bioconductor.org/biocLite.R")
biocLite("IRanges")
```

To cite this package in a publication, start R and enter:

```
citation("IRanges")
```

Documentation

[PDF](#) [R Scripts](#) An Introduction to IRanges
[PDF](#) [R Scripts](#) Rle Tips and Tricks
[PDF](#) Reference Manual
[Text](#) NEWS

Details

biocViews [DataRepresentation](#), [Infrastructure](#), [Software](#)
Version 1.22.6
In Bioconductor BioC 2.3 (R-2.8)
since

Workflows

Common Bioconductor workflows include:

- [Oligonucleotide Arrays](#)
- [High-throughput Sequencing](#)
- [Counting Reads for Differential Expression](#) (parathyroidSE vignette)
- [Annotation](#)
- [Annotating Variants](#)
- [Annotating Ranges](#)
- [Flow Cytometry](#) and other assays
- [Candidate Binding Sites for Known Transcription Factors](#)
- [Cloud-enabled cis-eQTL search and annotation](#)

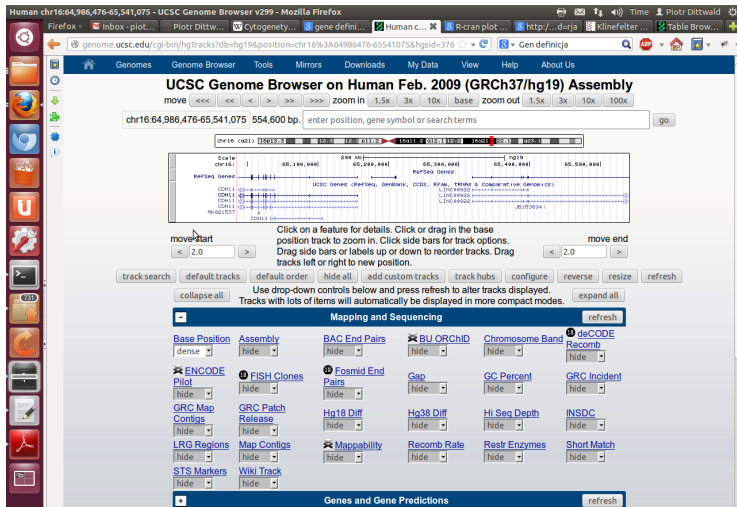
Mailing Lists

Post questions about Bioconductor packages to our mailing lists. Read the [posting guide](#) before posting!

- [bioconductor](#)
- [bioc-devel](#)

<http://genome.ucsc.edu/cgi-bin/hgGateway>

- ▶ zbiory danych:
 - ▶ do oglądania w przeglądarce UCSC
 - ▶ do ściągnięcia przez Table Browser
- ▶ np. RefSeq, UCSC Genes



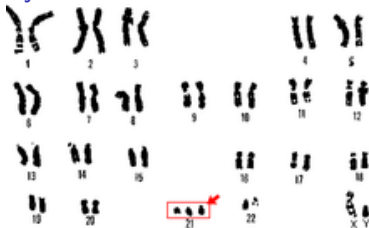
Aberracje chromosomowe

Syndrom Klinefeltera



źródło: wikipedia

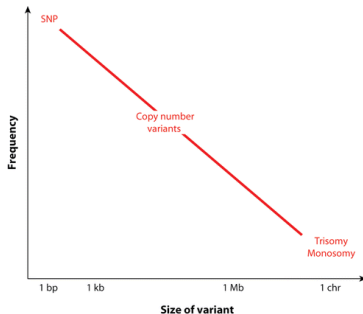
Syndrom Downa



źródło: wikipedia

CNVs

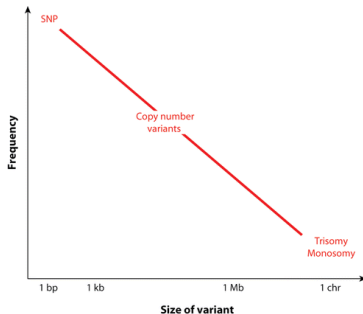
- ▶ obszary genomu, których u danego osobnika jest mniej lub więcej niż w referencyjnym genomie, nazywamy wariantami o zmienionej liczbie kopii (*ang. Copy-Number Variants; CNVs*)




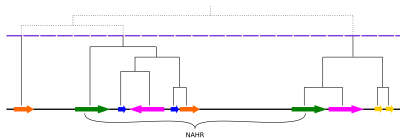
Girirajan S, et al. 2011.
Annu. Rev. Genet. 45:203–26

CNVs

- ▶ obszary genomu, których u danego osobnika jest mniej lub więcej niż w referencyjnym genomie, nazywamy wariantami o zmienionej liczbie kopii (*ang. Copy-Number Variants; CNVs*)



 Girirajan S, et al. 2011.
Annu. Rev. Genet. 45:203–26

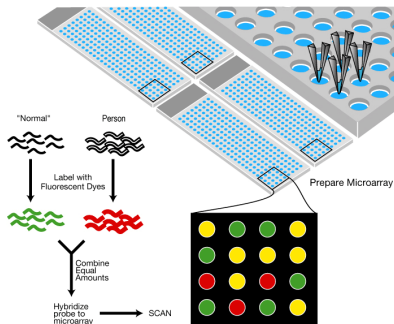


- ▶ w szczególności CNVs występują między długimi (10-400 kb) fragmentami DNA o wysokim (> 97%) współczynniku podobieństwa sekwencyjnego (mechanizm NAHR)

Całogenomowa mapa niestabilności - delecje i duplikacje



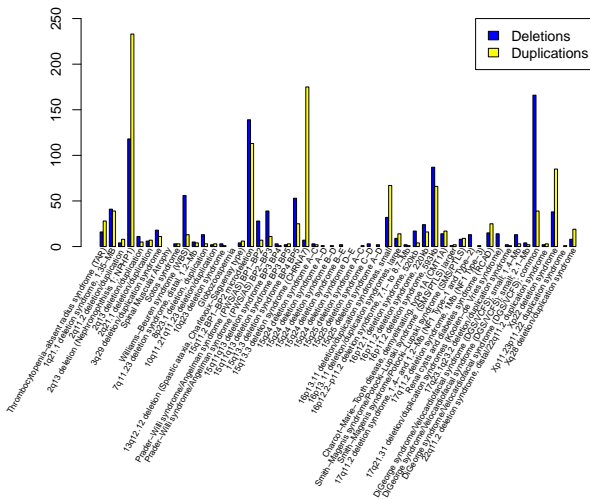
Dane kliniczne



źródło: atlantichealth.dnadirect.com

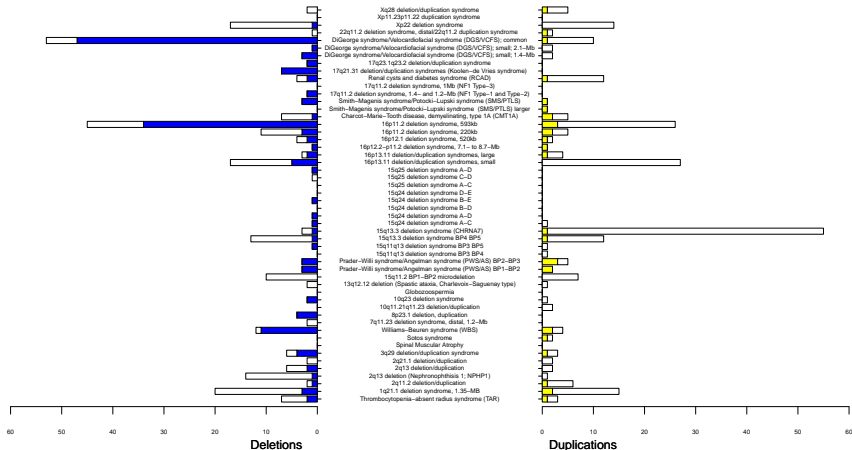
- ▶ metoda porównawczej hybrydyzacji genomowej do mikromacierzy (*ang. microarray-based Comparative Genomic Hybridization; aCGH*)
- ▶ dane kliniczne > 25,000 pacjentów z bazy Baylor College of Medicine, Houston

Dla których chorób jest najwięcej pacjentów?



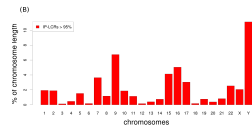
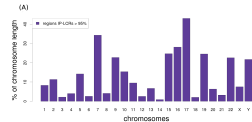
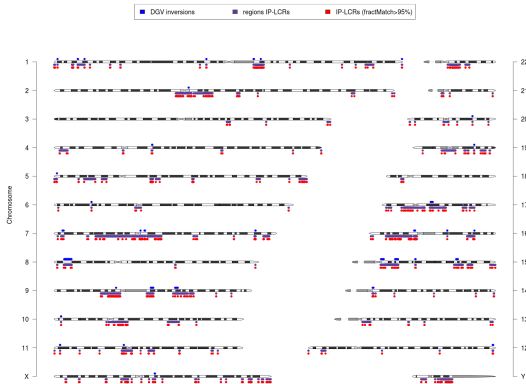
Częstość występowania (bez informacji o dziedziczeniu) znanych zespołów chorobowych związanych z nawracającymi delecjami i duplikacjami.

Gdzie występuje najczęściej rearanżacji *de novo*?



Pacjenci dla przypadków *de novo*

Całogenomowa mapa niestabilności - inwersje



(A) Obszar genomu (A) otoczony oraz (B) zajmowany przez jednostki mediujące nawracające inwersje.

Informacje przydatne w diagnostyce

Geny potencjalnie uszkodzone przez nawracające inwersje, a już skojarzone z jednostkami chorobowymi (fragment tabeli)

Gene	Gene description	Location	Intersection with LCR Size (Size of entire LCR) kb	LCR identity %	Disease	Inheritance	OMIM
<i>ABCC6</i>	ATP-binding cassette, sub-family C (CFTR/MRP), member 6	16p13.11	25 (128)	99.36	Pseudoxanthoma elasticum	AR	264800
<i>AKR1C2</i>	Aldo-keto reductase family 1, member C2	10p15.1	28 (47.5)	95.15	46,XY sex reversal 8	AR	614279
<i>BCR</i>	Breakpoint cluster region	22q11.23	7(10.5), 4(7)	95.98; 96.17	Chronic myeloid leukemia (CML)	-	608232

Podziękowania

- ▶ Tomasz Gambin (Politechnika Warszawska/obecnie BCM)
- ▶ Anna Gambin (Uniwersytet Warszawski)
- ▶ Paweł Stankiewicz, James R. Lupski, Chad A. Shaw, Sau Wai Cheung (BCM)
- ▶ UE przez EFS (UDA-POKL.04.01.01-00-072/09-00), NCN, Fundacja Nauki Polskiej (stypendium START)

Więcej

- ▶ Piotr Dittwald*, Tomasz Gambin* et al (2013). NAHR-mediated copy-number variants in a clinical population: Mechanistic insights into both genomic disorders and Mendelizing traits. (* wkład równorzędny) *Genome Research* 23, 9: 1395-409,
- ▶ Piotr Dittwald*, Tomasz Gambin*, Claudia Gonzaga-Jauregui* et al (2013). Inverted low-copy repeats and genome instability – a genome-wide analysis. (* wkład równorzędny) *Human Mutation* 34, 1: 210-20.

Research

NAHR-mediated copy-number variants in a clinical population: Mechanistic insights into both genomic disorders and Mendelizing traits

Piotr Dittwald,^{1,2,3,17} Tomasz Gambin,^{1,4,17} Przemysław Szafranski,¹ Jian Li,¹ Stephen Amato,⁵ Michael Y. Divon,⁶ Lisa Ximena Rodríguez Rojas,⁷ Lindsay E. Elton,⁸ Daryl A. Scott,^{1,9} Christian P. Schaaf,¹ Wilfredo Torres-Martínez,¹⁰ Abby K. Stevens,¹⁰ Jill A. Rosenfeld,¹¹ Satish Agadi,¹² David Francis,¹³ Sung-Hae L. Kang,¹ Amy Berman,¹ Seema R. Lalani,¹ Carlos A. Bacino,¹ Weimin Bi,¹ Aleksandar Milosavljevic,¹ Arthur L. Beaudet,¹ Ankita Patel,¹ Chad A. Shaw,¹ James R. Lupski,^{1,14,15} Anna Gambin,^{2,16} Sau Wai Cheung,¹ and Paweł Stankiewicz^{1,18}

RESEARCH ARTICLE

Human Mutation



Inverted Low-Copy Repeats and Genome Instability—A Genome-Wide Analysis

Piotr Dittwald,^{1,21} Tomasz Gambin,³⁺¹ Claudia Gonzaga-Jauregui,⁴¹ Claudia M.B. Carvalho,⁴ James R. Lupski,⁴⁻⁹ Paweł Stankiewicz,^{4,16} and Anna Gambin^{1,7}

¹Institute of Informatics, University of Warsaw, Warsaw, Poland; ²College of Inter-Faculty Individual Studies in Mathematics and Natural Sciences, University of Warsaw, Warsaw, Poland; ³Institute of Computer Science, Warsaw University of Technology, Warsaw, Poland;

⁴Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas; ⁵Pediatrics, Baylor College of Medicine, Houston, Texas; ⁶Texas Children's Hospital, Houston, Texas; ⁷Mossakowski Medical Research Centre, Polish Academy of Sciences, Warsaw, Poland

Communicated by David N. Cooper

Received 5 June 2012; accepted revised manuscript 30 August 2012

Published online 10 September 2012 in Wiley Online Library (www.interscience.wiley.com/jpages/hummutat). DOI: 10.1002/humut.22277

Dziękuję za uwagę!

Dziękuję za uwagę!

Ciekawy projekt? Proszę o kontakt:
piotr.dittwald@gmail.com