

Introduction

Measuring
FCGR3 CNVs

DOC-based
methods?
A DOC-based
pipeline

LD between
CNVs and SNPs
Method
Results
Conclusion

Incorporating
PEM into the
pipeline
PEM?
Breakpoints with
PEM?

Next stage

Acknowledgments

References

Is copy number variation at the FCGR locus able to be tagged by single nucleotide polymorphisms?

Hoang Tan Nguyen

Tony Merriman

Mik Black

Department of Mathematics and Statistics

Department of Biochemistry

The University of Otago

Presentation

Introduction

Measuring
FCGR3 CNVs

DOC-based
methods?
A DOC-based
pipeline

LD between
CNVs and SNPs
Method
Results
Conclusion

Incorporating
PEM into the
pipeline
PEM?
Breakpoints with
PEM?

Next stage

Acknowledgments

References

- ① **Introduction.**
- ② **Measuring Copy Number Variation (CNV) at the FCGR3 locus**
- ③ **Calculating linkage disequilibrium (LD) between CNVs and nearby single nucleotide polymorphisms (SNPs).**
- ④ **Detecting breakpoints in simulated samples having FCGR3A/3B genes duplicated.**

Introduction

Measuring FCGR3 CNVs

DOC-based
methods?
A DOC-based
pipeline

LD between
CNVs and SNPs
Method
Results
Conclusion

Incorporating
PEM into the
pipeline
PEM?
Breakpoints with
PEM?

Next stage

Acknowledgments

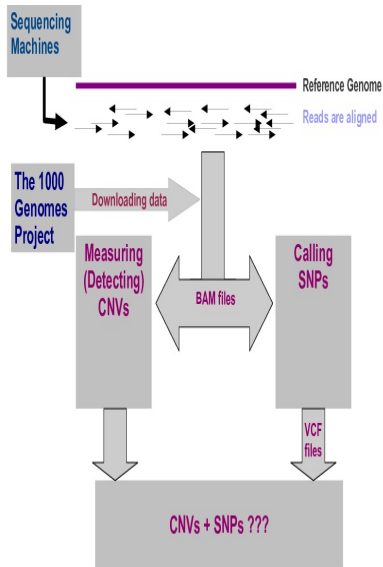
References

Use sequencing data:

- 1 Measuring CNVs at complex loci conferring susceptibility to diseases.
- 2 Predicting tagSNPs for the loci.

At this stage, we are using data from the 1000 Genomes Project and simulated data.

Our group's work



The FCGR3 locus?

FCGR3A/3B CNVs are risk factors for disease and have been investigated intensively.

Introduction

Measuring

FCGR3 CNVs

DOC-based
methods?

A DOC-based
pipeline

LD between

CNVs and SNPs

Method

Results

Conclusion

Incorporating

PEM into the
pipeline

PEM?

Breakpoints with
PEM?

Next stage

Acknowledgments

References

Introduction

Measuring
FCGR3 CNVs

DOC-based
methods?
A DOC-based
pipeline

LD between
CNVs and SNPs

Method
Results
Conclusion

Incorporating
PEM into the
pipeline

PEM?

Breakpoints with
PEM?

Next stage

Acknowledgments

References

The FCGR3 locus?

FCGR3A/3B CNVs are risk factors for disease and have been investigated intensively.

BRIEF DEFINITIVE REPORT

FCGR3B copy number variation is associated with susceptibility to systemic, but not organ-specific, autoimmunity

Manuela Fanciulli^{1,11}, Penny J Norsworthy^{1,11}, Enrico Petretto^{1,11}, Rong Dong¹, Lorraine Harper², Lavanya Kamesh², Joanne M Heward³, Stephen C L Gough³, Adam de Smith⁴, Alexandra I F Blakemore⁴, Philippe Froguel^{4,5}, Catherine J Owen⁶, Simon H S Pearce⁶, Luis Teixeira⁷, Loïc Guillevin⁷, Deborah S Cunningham-Graham⁸, Charles D Pusey⁹, H Terence Cook¹⁰, Timothy J Vyse⁸ & Timothy J Aitman^{1,8}

Ann Rheum Dis. 2010 Sep;69(9):1711-6. Epub 2010 May 14.

Association of variation in Fcγ receptor 3B gene copy number with rheumatoid arthritis in Caucasian samples.

McKinney C, Fanciulli M, Merriman ME, Phipps-Green A, Alizadeh BZ, Koeleman BP, Dalbeth N, Gow PJ, Harrison AA, Highton J, Jones PB, Stamp LK, Steer S, Barrera P, Coenen MJ, Franke B, van Riel PL, Vyse TJ, Aitman TJ, Radstake TR, Merriman TR.

Department of Biochemistry, University of Otago, Dunedin, New Zealand.

Copy number of *FCGR3B*, which is associated with systemic lupus erythematosus, correlates with protein expression and immune complex uptake

Lisa C. Willcocks,^{1,2} Paul A. Lyons,^{1,2} Menna R. Clatworthy,^{1,2} James I. Robinson,⁴ Wanling Yang,⁵ Stephen A. Newland,^{1,2} Vincent Plagnol,^{1,3} Naomi N. McGovern,² Alison M. Condliffe,² Edwin R. Chilvers,² Dwomoa Adu,⁶ Elaine C. Jolly,^{1,2} Richard Watts,⁷ Yu Lung Lau,⁵ Ann W. Morgan,⁴ Gerard Nash,⁸ and Kenneth G.C. Smith^{1,2}

Copy number polymorphism in *Fcgr3* predisposes to glomerulonephritis in rats and humans

Timothy J. Aitman¹, Rong Dong^{1*}, Timothy J. Vyse^{2*}, Penny J. Norsworthy^{1*}, Michelle D. Johnson¹, Jennifer Smith³, Jonathan Mangion¹, Cheri Robertson-Lowe^{1,2}, Amy J. Marshall¹, Enrico Petretto¹, Matthew D. Hodges¹, Gurjeet Bhargal³, Sheetal G. Patel¹, Kelly Sheehan-Rooney¹, Mark Duda^{1,3}, Paul R. Cook^{1,3}



Introduction

Measuring
FCGR3 CNVs

DOC-based
methods?
A DOC-based
pipeline

LD between
CNVs and SNPs

Method
Results
Conclusion

Incorporating
PEM into the
pipeline

PEM?
Breakpoints with
PEM?

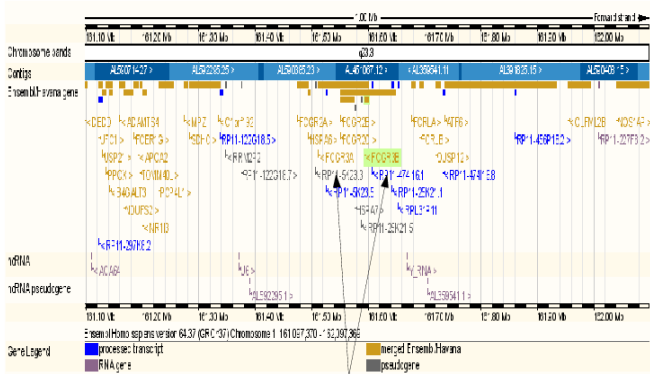
Next stage

Acknowledgments

References

The FCGR3 locus?

Measuring and genotyping at the FCGR3 locus is complex in view of **the high level of structural homology** between FCGR3A and FCGR3B (Morgan et al., 2006; Hollox et al., 2009).



The two genes FCGR3A and FCGR3B (GRCh37)

Introduction

Measuring
FCGR3 CNVs

DOC-based
methods?
A DOC-based
pipeline

LD between
CNVs and SNPs

Method
Results
Conclusion

Incorporating
PEM into the
pipeline

PEM?
Breakpoints with
PEM?

Next stage

Acknowledgments

References

Can we use a set of SNPs
to predict the CNV status at the FCGR locus?

Introduction

Measuring
FCGR3 CNVs

DOC-based
methods?
A DOC-based
pipeline

LD between
CNVs and SNPs

Method
Results
Conclusion

Incorporating
PEM into the
pipeline

PEM?
Breakpoints with
PEM?

Next stage

Acknowledgments

References

Can we use a set of SNPs to predict the CNV status at the FCGR locus?

We:

- ① Developed a pipeline to measure FCGR3A/3B CNVs.
- ② Calculated LD between SNPs and FCGR3A/3B CNVs.
- ③ Are developing a method to predict breakpoints.

Depth of coverage (DOC) based method

We developed and used DOC-based software to measure CNVs at the FCGR3 locus.

Introduction

Measuring
FCGR3 CNVs

DOC-based
methods?

A DOC-based
pipeline

LD between
CNVs and SNPs

Method

Results

Conclusion

Incorporating
PEM into the
pipeline

PEM?

Breakpoints with
PEM?

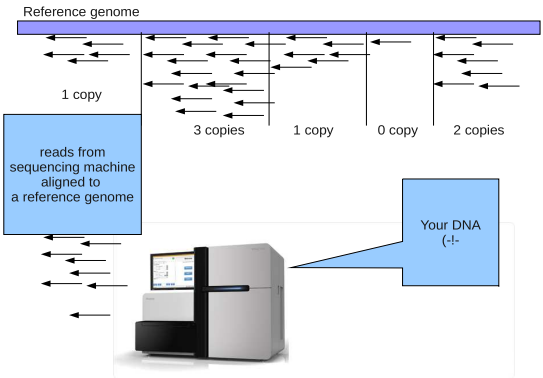
Next stage

Acknowledgments

References

Depth of coverage (DOC) based method

We developed and used DOC-based software to measure CNVs at the FCGR3 locus.



A DOC-based pipeline

- Use **CNVnator** (Abyzov et al., 2011) + **cnvMB**¹.

Introduction

Measuring

FCGR3 CNVs

DOC-based
methods?

A DOC-based
pipeline

LD between
CNVs and SNPs

Method

Results

Conclusion

Incorporating
PEM into the
pipeline

PEM?

Breakpoints with
PEM?

Next stage

Acknowledgments

References

¹In house tool

A DOC-based pipeline

- Use **CNVnator** (Abyzov et al., 2011) + **cnvMB**¹.
- Compare with results published (Hollox et al., 2009) to get reliable **thresholds** for **cnvMB**.

Introduction

Measuring
FCGR3 CNVs

DOC-based
methods?

A DOC-based
pipeline

LD between
CNVs and SNPs

Method

Results

Conclusion

Incorporating
PEM into the
pipeline

PEM?

Breakpoints with
PEM?

Next stage

Acknowledgments

References

¹In house tool

A DOC-based pipeline

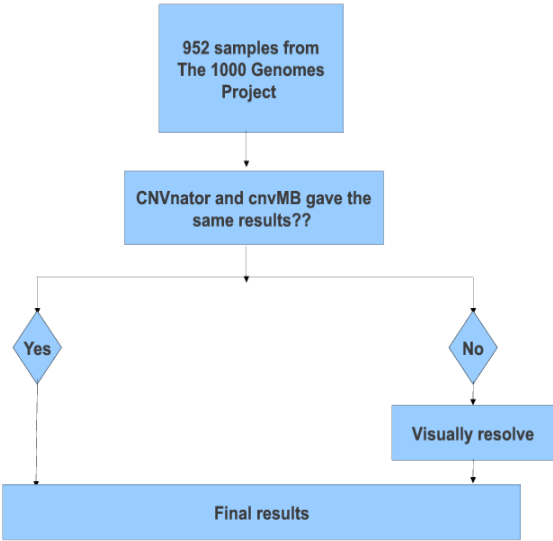
- Use **CNVnator** (Abyzov et al., 2011) + **cnvMB**¹.
- Compare with results published (Hollox et al., 2009) to get reliable **thresholds** for **cnvMB**.
-

The screenshot shows the top half of a research paper page on the 'GENOME RESEARCH' website. The paper title is 'CNVnator: An approach to discover, genotype and characterize typical and atypical CNVs from family and population genome sequencing'. The authors listed are Alexey Abyzov¹, Alexander Eckehart Urban², Michael Snyder², and Mark Gerstein^{1,3}. The paper is marked as an 'ACCEPTED PREPRINT'. On the right, there is a 'Current Issue' section for May 2011, 24 (3), and a 'From the Cover' section. Below the paper title, a 'METHODS' section is visible, followed by the title 'An Integrated Approach for Measuring Copy Number Variation at the FCGR3 (CD16) Locus' by Edward J. Hollox, Jan-Christoph Detering, and Tushna Behnugara. The authors' affiliations and the paper's publication details (received 1 July 2008, accepted 28 August 2008, published online 13 January 2009) are also shown.

- Resolve all discordant samples by visual inspection.

¹In house tool

A DOC-based pipeline



Measuring the LD between CNVs and SNPs

Introduction

Measuring
FCGR3 CNVs

DOC-based
methods?
A DOC-based
pipeline

LD between
CNVs and SNPs

Method
Results
Conclusion

Incorporating
PEM into the
pipeline
PEM?
Breakpoints with
PEM?

Next stage

Acknowledgments

References

- ① Extract SNPs in a 2MB region flanking the genes => 25878 SNPs in and nearby the loci.
- ② Fisher's exact test.
- ③ Recheck visually samples having tag SNPs ($p - \text{adjusted} < 0.05$).
- ④ *FCGR3B* CNVs and some SNPs are in LD in some populations.

SNPs in LD with duplication of FCGR3B

Introduction

Measuring
FCGR3 CNVs

DOC-based
methods?
A DOC-based
pipeline

LD between
CNVs and SNPs

Method

Results

Conclusion

Incorporating
PEM into the
pipeline

PEM?

Breakpoints with
PEM?

Next stage

Acknowledgments

References

- 1 Dividing duplication at the FCGR3A/3B locus into four classes: duplication at the FCGR3A (A), FCGR3B (B), both (AB) and normal.

SNPs in LD with duplication of FCGR3B

Introduction

Measuring
FCGR3 CNVs

DOC-based
methods?
A DOC-based
pipeline

LD between
CNVs and SNPs

Method
Results
Conclusion

Incorporating
PEM into the
pipeline
PEM?
Breakpoints with
PEM?

Next stage

Acknowledgments

References

- 1 Dividing duplication at the FCGR3A/3B locus into four classes: duplication at the FCGR3A (**A**), FCGR3B (**B**), both (**AB**) and **normal**.
- 2 Tag SNPs were found in Mexican ancestry in Los Angeles, California (**MXL**), Japanese (**JPT**), Chinese (**CHS**, **CHB**) samples sets.

SNPs in LD with duplication of FCGR3B

Introduction

Measuring
FCGR3 CNVs

DOC-based
methods?
A DOC-based
pipeline

LD between
CNVs and SNPs

Method
Results
Conclusion

Incorporating
PEM into the
pipeline
PEM?
Breakpoints with
PEM?

Next stage

Acknowledgments

References

- 1 Dividing duplication at the FCGR3A/3B locus into four classes: duplication at the FCGR3A (**A**), FCGR3B (**B**), both (**AB**) and **normal**.
- 2 Tag SNPs were found in Mexican ancestry in Los Angeles, California (**MXL**), Japanese (**JPT**), Chinese (**CHS**, **CHB**) samples sets.

The **most significant results** in the **MXL** population (*described by the percentage of samples having SNPs*):

| SNP (position) | A (n = 4) | B (n = 7) | AB (n =6) | Normal (n = 32) | p.values |
|------------------|-----------|-----------|-----------|-----------------|----------|
| 161530185 | 0 | 100 | 100 | 3.1 | 1.4E-007 |
| 161610869 | 0 | 100 | 100 | 3.1 | 1.4E-007 |
| 161610920 | 0 | 100 | 100 | 3.1 | 1.4E-007 |
| 161616926 | 0 | 100 | 100 | 3.1 | 1.4E-007 |
| 161617022 | 0 | 100 | 100 | 3.1 | 1.4E-007 |
| 161528550 | 0 | 100 | 100 | 6.3 | 7.0E-007 |

Bold SNPs are good tagSNPs in the other three populations.

Conclusion

Introduction

Measuring
FCGR3 CNVs

DOC-based
methods?
A DOC-based
pipeline

LD between
CNVs and SNPs

Method
Results
Conclusion

Incorporating
PEM into the
pipeline
PEM?
Breakpoints with
PEM?

Next stage

Acknowledgments

References

- It is possible to use SNPs to tag copy number variation at the FCGR3B locus, but copy-number and SNPs have to be characterized on a population-specific basis.
- Deletion tag SNPs were not detected in other African and Causasion samples
=> false positive or extremely population specific.

Conclusion

Introduction

Measuring
FCGR3 CNVs

DOC-based
methods?
A DOC-based
pipeline

LD between
CNVs and SNPs

Method
Results
Conclusion

Incorporating
PEM into the
pipeline
PEM?
Breakpoints with
PEM?

Next stage

Acknowledgments

References

- It is possible to use SNPs to tag copy number variation at the FCGR3B locus, but copy-number and SNPs have to be characterized on a population-specific basis.
- Deletion tag SNPs were not detected in other African and Causasion samples
=> false positive or extremely population specific.

To know more about localized structure, bases immediately adjacent to breakpoints can potentially be assayed to infer CNV.

Conclusion

- It is possible to use SNPs to tag copy number variation at the FCGR3B locus, but copy-number and SNPs have to be characterized on a population-specific basis.
- Deletion tag SNPs were not detected in other African and Causasion samples
=> false positive or extremely population specific.

To know more about localized structure, bases immediately adjacent to breakpoints can potentially be assayed to infer CNV.

=> We decided to incorporate **paired-end read based method** into the pipeline to predict **breakpoints** at the FCGR locus of samples having duplicated FCGR3A/3B genes.

Some limitations of DOC

Introduction

Measuring

FCGR3 CNVs

DOC-based
methods?

A DOC-based
pipeline

LD between
CNVs and SNPs

Method

Results

Conclusion

Incorporating
PEM into the
pipeline

PEM?

Breakpoints with
PEM?

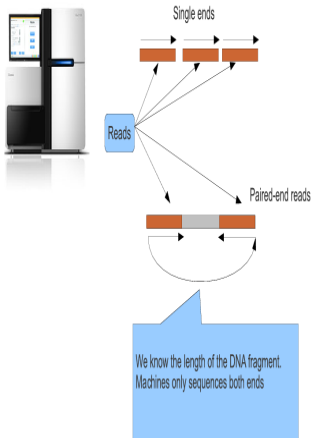
Next stage

Acknowledgments

References

- ❶ Cannot pinpoint breakpoints exactly.
- ❷ Gives ambiguous results at repetitive regions.
=> We are using a paired-end reads based method (PEM) to improve.
- ❸ Genotyping the exact number of CNV is still challenging.
=> In the future, we will combine with experimental methods to genotype more precisely

Paired-end sequencing



Paired-end read mapping (PEM) based methods have been used by *Ye, K. et al. (2009)*; *Hormozdiari et al. (2009)*; *Chen et al. (2009)*; *Mills et al. (2011)*; *Hormozdiari et al. (2011)* to detect structural variations:

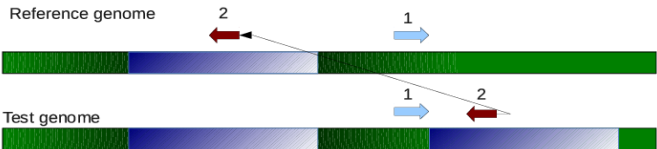
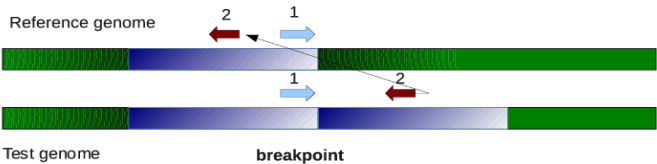
- Deletion.
- Inversion.
- Insertion.

PEM

Find out duplication and their potential breakpoints by **clustering** paired reads **whose orientations are not as expected**.

Example: Tandem duplication and duplication + translocation.

Duplication (Haploid case)



PEM with simulated data

- Simulate duplication³ at the FCGR3 locus with different kinds:
Only 3A.
Only 3B.
Segmental duplication at 3A and 3B.

Introduction

Measuring

FCGR3 CNVs

DOC-based
methods?

A DOC-based
pipeline

LD between
CNVs and SNPs

Method

Results

Conclusion

Incorporating
PEM into the
pipeline

PEM?

Breakpoints with
PEM?

Next stage

Acknowledgments

References

³Use *dwgsim*: <http://sourceforge.net/apps/mediawiki/dnaa>

PEM with simulated data

Introduction

Measuring

FCGR3 CNVs

DOC-based
methods?

A DOC-based
pipeline

LD between
CNVs and SNPs

Method

Results

Conclusion

Incorporating
PEM into the
pipeline

PEM?

Breakpoints with
PEM?

Next stage

Acknowledgments

References

- Simulate duplication³ at the FCGR3 locus with different kinds:
Only 3A.
Only 3B.
Segmental duplication at 3A and 3B.
- Cluster discordant paired ends and predict potential breakpoints.

³Use *dwgsim*: <http://sourceforge.net/apps/mediawiki/dnaa>

PEM with simulated data

Introduction

Measuring

FCGR3 CNVs

DOC-based
methods?

A DOC-based
pipeline

LD between
CNVs and SNPs

Method

Results

Conclusion

Incorporating
PEM into the
pipeline

PEM?

Breakpoints with
PEM?

Next stage

Acknowledgments

References

- Simulate duplication³ at the FCGR3 locus with different kinds:
Only 3A.
Only 3B.
Segmental duplication at 3A and 3B.
- Cluster discordant paired ends and predict potential breakpoints.
- Compare with **Breakdancer** (Chen et al., 2009)

Nature Methods **6**, 677 - 681 (2009)

Published online: 9 August 2009 | doi:10.1038/nmeth.1363

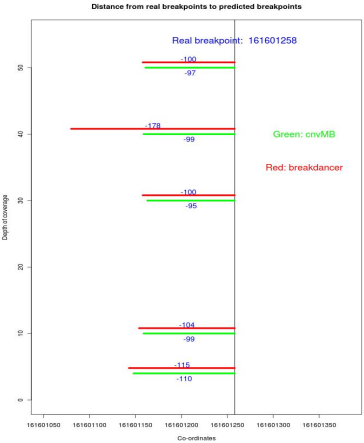
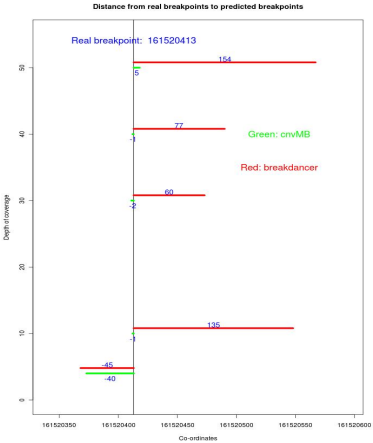
BreakDancer: an algorithm for high-resolution mapping of genomic structural variation

Ken Chen¹, John W Wallis¹, Michael D McLellan¹, David E Larson¹, Joelle M Kalicki¹,
Craig S Pohl¹, Sean D McGrath¹, Michael C Wendl¹, Qunyuan Zhang², Devin P Locke¹,
Xiaoqi Shi¹, Robert S Fulton¹, Timothy J Ley¹, Richard K Wilson¹, Li Ding¹ & Elaine R
Mardis¹

³Use *dwgsim*: <http://sourceforge.net/apps/mediawiki/dnaa>

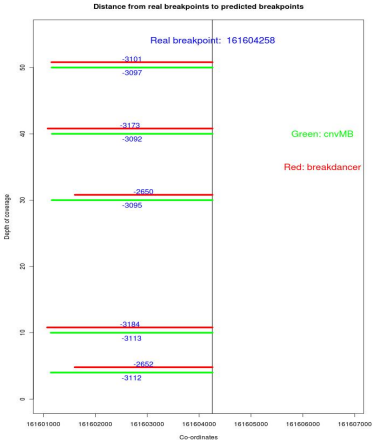
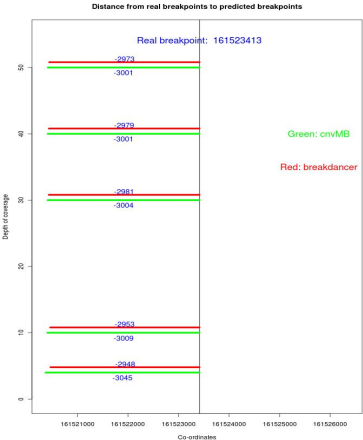
PEM with simulated data

Both **cnvMB** and **Breakdancer** worked well with tandem duplication.



PEM with simulated data

- Both **cnvMB** and **Breakdancer**⁴ failed with duplication + translocation.



⁴Breakdancer has been designed to work with tandem duplication

Next stage

Introduction

Measuring

FCGR3 CNVs

DOC-based
methods?

A DOC-based
pipeline

LD between
CNVs and SNPs

Method

Results

Conclusion

Incorporating
PEM into the
pipeline

PEM?

Breakpoints with
PEM?

Next stage

Acknowledgments

References

The pipeline is being developed to

- Predict breakpoints for translocation + duplication.
- Incorporate mapping quality of reads into a probability frame.

After that, we will apply the pipeline to real data.

Acknowledgments

I am very grateful the **Virtual Institute of Statistical Genetics (VISG)** for the financial assistance.

Thank you very much

Introduction

Measuring
FCGR3 CNVs

DOC-based
methods?
A DOC-based
pipeline

LD between
CNVs and SNPs

Method

Results

Conclusion

Incorporating
PEM into the
pipeline

PEM?

Breakpoints with
PEM?

Next stage

Acknowledgments

References

Abyzov, A., Urban, A., Snyder, M., Gerstein, M., 2011. Cnvnator: An approach to discover, genotype and characterize typical and atypical cnvs from family and population genome sequencing. *Genome Research*.

Chen, K., Wallis, J., McLellan, M., Larson, D., Kalicki, J., Pohl, C., McGrath, S., Wendl, M., Zhang, Q., Locke, D., et al., 2009. Breakdancer: an algorithm for high-resolution mapping of genomic structural variation. *Nature Methods* 6 (9), 677–681.

Hollox, E., Detering, J., Dehnugara, T., 2009. An integrated approach for measuring copy number variation at the fcgr3 (cd16) locus. *Human mutation* 30 (3), 477–484.

Hormozdiari, F., Alkan, C., Eichler, E., Sahinalp, S., 2009. Combinatorial algorithms for structural variation detection in high-throughput sequenced genomes. *Genome research* 19 (7), 1270.

Hormozdiari, F., Hajirasouliha, I., McPherson, A., Eichler, E., Sahinalp, S., 2011. Simultaneous structural variation discovery in multiple paired-end sequenced genomes. In: *Research in Computational Molecular Biology*. Springer, pp. 104–105.

Mills, R., Walter, K., Stewart, C., Handsaker, R., Chen, K., Alkan, C., Abyzov, A., Yoon, S., Ye, K., Cheetham, R., et al., 2011. Mapping copy number variation by population-scale genome sequencing. *Nature* 470 (7332), 59–65.

Morgan, A., Robinson, J., Barrett, J., Martin, J., Walker, A., Babbage, S., Ollier, W., Gonzalez-Gay, M., Isaacs, J., 2006. Association of fcgr2a and fcgr2a-fcgr3a haplotypes with susceptibility to giant cell arteritis. *Arthritis Research & Therapy* 8 (4), R109.

Ye, K., Schulz, M. H., Long, Q., Apweiler, R., Ning, Z., 06 2009. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics* 25 (21), 2865–2871.