SV calling

PLOS | GENETICS

RESEARCH ARTICLE

Whole Genome Sequencing Identifies a 78 kb Insertion from Chromosome 8 as the Cause of Charcot-Marie-Tooth Neuropathy CMTX3

Megan H. Brewer[1,2]*, Rabia Chaudhry[1,2], Jessica Qi[1,3], Aditi Kidambi[1], Alexander P. Drew[1], Manoj P. Menezes[4,5,6], Monique M. Ryan[7,8,9], Michelle A. Farrar[10,11], David Mowat[11,12], Gopinath M. Subramanian[13], Helen K. Young[14,15,16], Stephan Zuchner[17,18], Stephen W. Reddel[19], Garth A. Nicholson[1,2,20], Marina L. Kennerson[1,2,20]

Phenotypic impact of genomic structural variation: insights from and for human disease

Joachim Weischenfeldt[1], Orsolya Symmons[2]*, François Spitz[2] and Jan O. Korbel[1]*

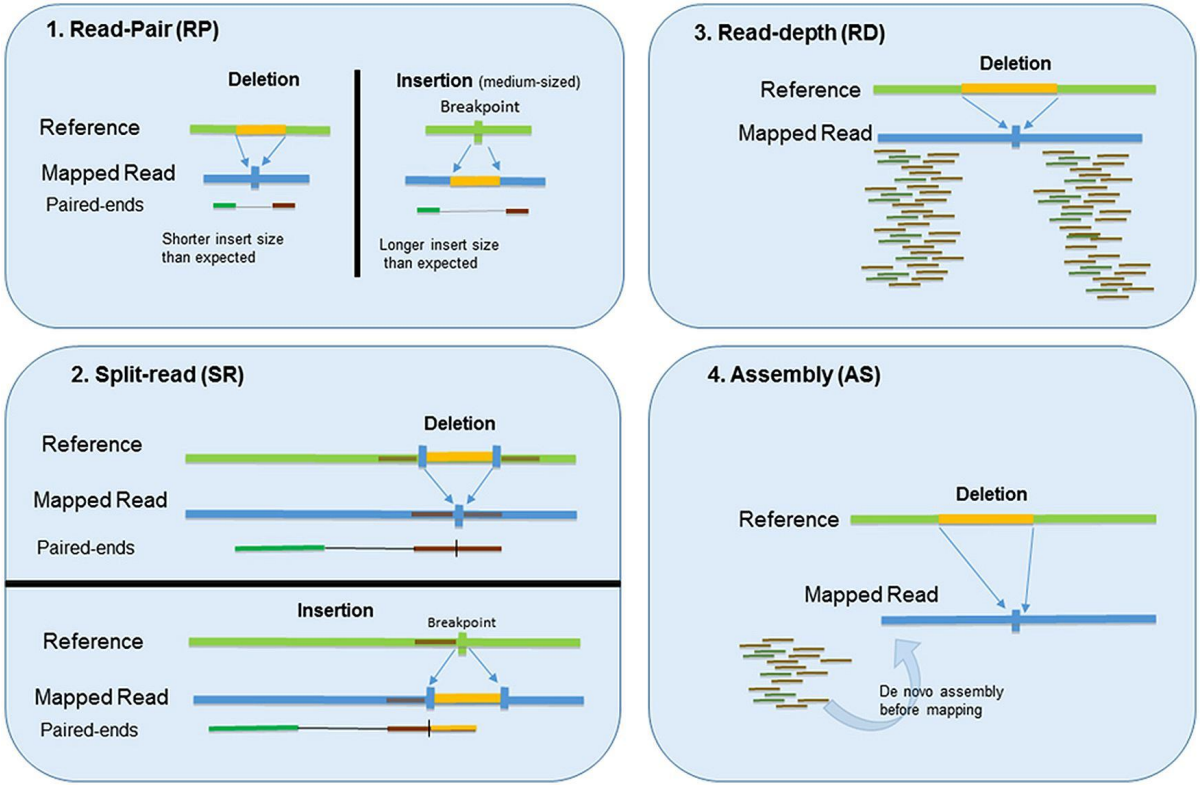
Copy number variations of chromosome 16p13.1 region associated with schizophrenia

A Ingason[1,2], D Rujescu[3], S Cichon[4,5], E Sigurdsson[6], T Sigmundsson[6], OPH Pietiläinen[7], JE Buizer-Voskamp[8,9], E Strengman[9], C Francks[10], P Muglia[10], A Gylfason[1], O Gustafsson[1], PI Olason[1], S Steinberg[1], T Hansen[2], KD Jakobsen[2], HB Rasmussen[2], I Giegling[3], H-J Möller[3], A Hartmann[3], C Crombie[11], G Fraser[11], N Walker[12], J Lonnqvist[13], J Suvisaari[13], A Tuulio-Henriksson[13], E Bramon[14], LA Kiemeney[15], B Franke[16], R Murray[14], E Vassos[14], T Toulopoulou[14], TW Mühleisen[4], S Tosato[17], M Ruggeri[17], S Djurovic[18,19], OA Andreassen[18,19], Z Zhang[20], T Werge[2], RA Ophoff[9,21], GROUP Investigators[22], M Rietschel[23], MM Nöthen[4,5], H Petursson[6], H Stefansson[1], L Peltonen[7,24,25], D Collier[14], K Stefansson[1], and DM St Clair[11]

Copy number variation of E3 ubiquitin ligase genes in peripheral blood leukocyte and colorectal cancer

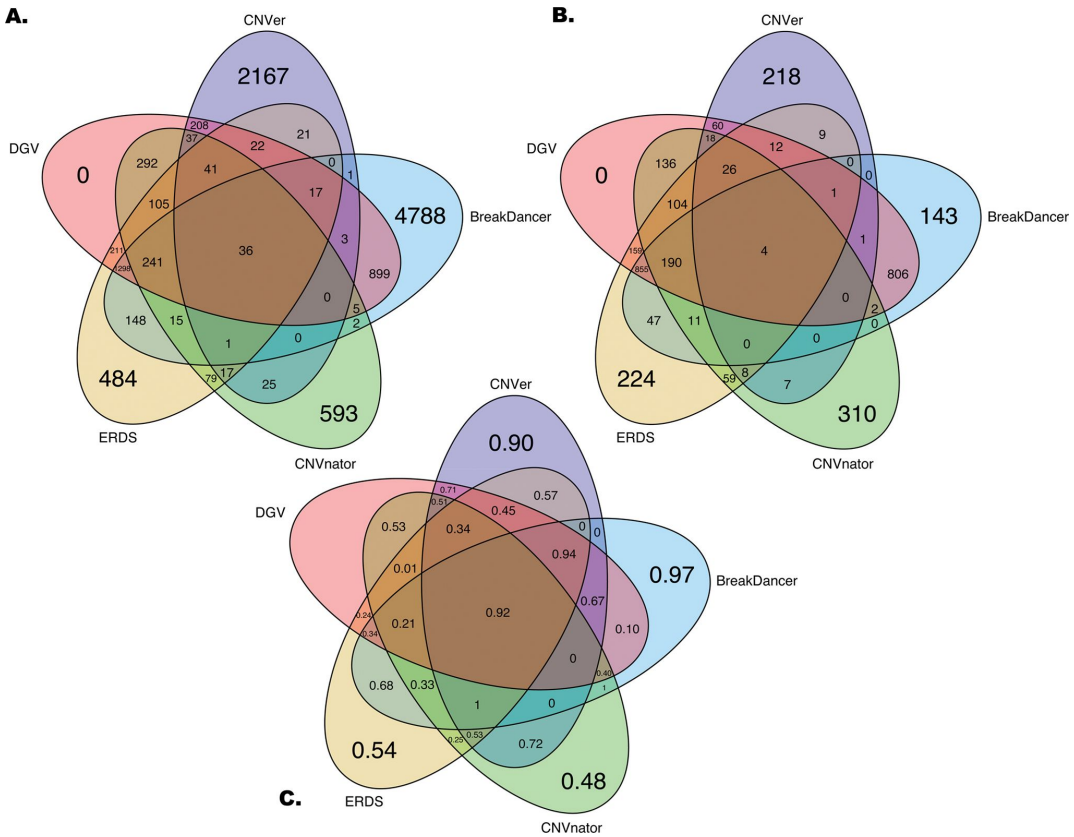
Haoran Bi[1], Tian Tian[1], Lin Zhu[1], Haibo Zhou[1], Hanqing Hu[2], Yanhong Liu[3], Xia Li[4], Fulan Hu[1], Yashuang Zhao[1] & Guiyu Wang[2]

SV short read evidence types



http://dx.doi.org/10.3389/fgene.2015.00138

The more the merrier?



DOI:10.1371/journal.pone.0122287

The first diagram (A.) represents the mean number of CNVs shared between tools prior to any filtering based on familial relationship. B. is the mean number of CNVs shared between the tools after filtering for Mendelian inheritance (*i.e.* CNVs that are in both twins and at least one parent). C. is the ratio of lost CNVs when filtering for Mendelian inheritance (). DGV is the Database of Genomic Variants.

1000 Genomes phase 3 SVs

Table 1: Phase 3 extended SV release

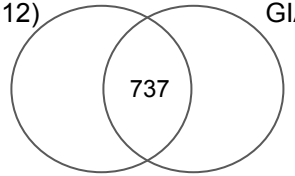
SV class	No. sites	Median size of SV sites (bp)	Median kbp per individual	Median alleles per individual	Site FDR	Biallelic site breakpoint precision (bp)	Genotype concordance (non-ref.)	Sensitivity estimates
Deletion (biallelic)	42,279	2,455	5,615	2,788	2%*–4%†	15 (±50)** 0.7 (±9.5)††	98%¶	88%¶
Duplication (biallelic)	6,025	35,890	518	17	1%*–4%†	683 (±1,350)‡‡	94%¶	65%¶
mCNV	2,929	19,466	11,346	340	1%*–4%†	–	NA	NA
Inversion	786	1,697	78	37	17%§ (9%)‡‖‖	32 (±47)‖‖	96%§	32%
MEI	16,631	297	691	1,218	4%‡	0.95 (±5.93)	98%‖	83# –96%•
NUMT	168	157	3	5.3	10%‡	0.25 (±0.43)	86.1%‡	NA

Nature 526, 75–81 doi:10.1038/nature15394

SV caller benchmark (GIAB NA12878 DEL set)

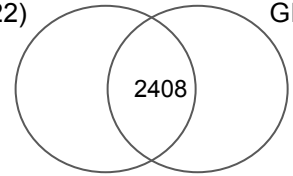
In the framework of the SciLifelab LTS ToolBox we benchmarked a number of callers. A combination of Manta only and Manta & Fermikit yielded the best sensitivity and precision, respectively

Lumpy (N=1912)
P=0.39



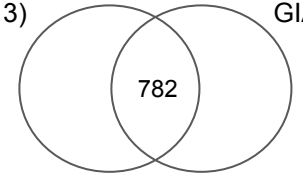
GIAB (N=2676)

Manta (N=5522)
P=0.44



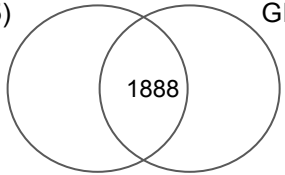
GIAB (N=2676)

Delly (N=9113)
P=0.09



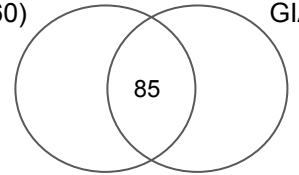
GIAB (N=2676)

Fermikit (N=3085)
P=0.61



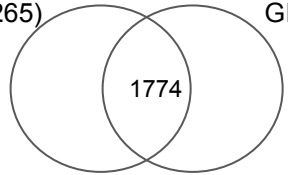
GIAB (N=2676)

Pindel (N=17560)
P=0.005



GIAB (N=2676)

Fermikit && Manta (N=2265)
P=0.78



GIAB (N=2676)

TIDDIT

SV caller developed @scilifelab (J. Eisfeldt, F. Vezzi, D. Nilsson et al. submitted)

Uses discordant pairs, read depth & split reads

Includes database feature for filtering

Detection of 822 deletions on NA12878

Caller	Sensitivity	Precision
CNVnator500	0.36	0.2
CNVnator200	0.18	0.06
Delly	0.91(0.85)	0.2(0.54)
FermiKit	0.83(0.77)	0.58(0.65)
TIDDIT	0.91(0.88)	0.67(**0.73**)
Lumpy	**0.92**	0.32
Manta	0.9(0.87)	0.66(0.72)

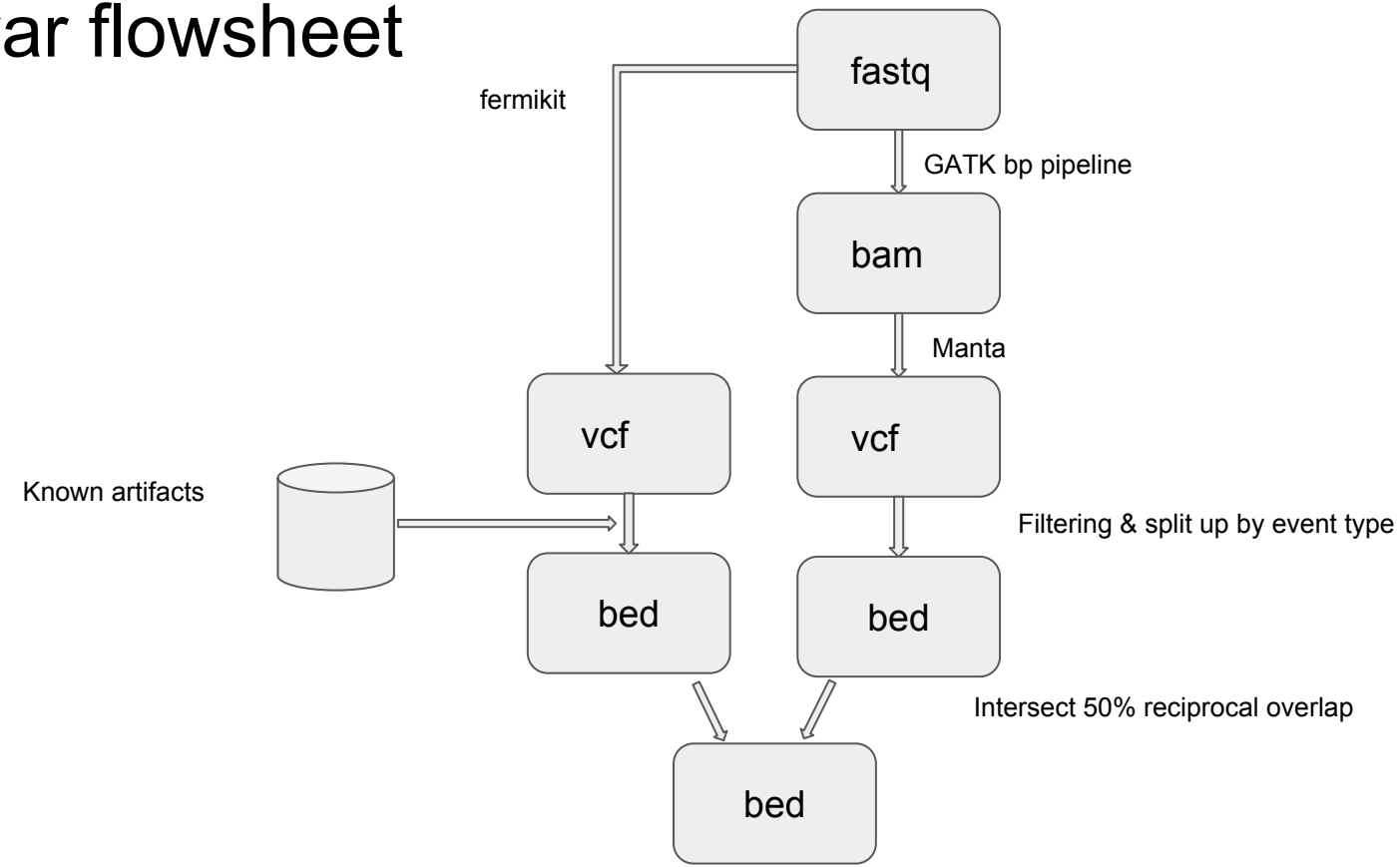
Removing “dubious” regions

LCR: 2% of the genome | 60-90% of calling errors (Li 2014. doi: 10.1093/bioinformatics/btu356)

Lumpy exclusion regions: Abnormal coverage in Illumina sequencing, probably assembly errors in b37 (Layer 2014. **DOI:** 10.1186/gb-2014-15-6-r84).

Moving to b38 will alleviate some problems

wgs-structvar flowsheet



Team Lean

The owner

The coder

The reviewer

The leader

“Harnessing change”

“Adjusting accordingly”

“Maximizing the amount of work NOT done”

Implementation

Best practices workflow ported from make to nextflow

Nextflow:

- Data driven pipelines
- Portable, parallel, reproducible

Other SciLifeLab nextflow projects:

- https://github.com/SciLifeLab/CAW : somatic variation analysis
- https://github.com/ewels/NGI-RNAseq : RNA seq best practice

This repository Search **Pull requests** **Issues** **Gist**


NBISweden / **wgs-structvar**

Unwatch 19 Star 1 Fork 2

Code Issues 2 Pull requests 2 Wiki Pulse Graphs

Whole Genome Sequenceing Structural Variation Pipelines

56 commits **4** branches **0** releases **4** contributors

Branch: **master** **New pull request** **Create new file** **Upload files** **Find file** **Clone or download**

samuell committed on **GitHub** Merge pull request #27 from NBISweden/feature/local-executor ... Latest commit `c01cfc7` 2 days ago

data	added small GIAB test bam file	3 months ago
docs	initial commit of empty data file makefile and flowsheet	3 months ago
src	Moved main.nf to base dir, should now be runnable without checkout	2 months ago
.gitignore	.gitignore .nextflow.* files	3 months ago
NOTES.md	Notes from standup	3 months ago
README.md	Update README.md	3 days ago
SVvcf2bed.pl	Adding the SVvcf2bed.pl script from pall	13 days ago
main.nf	Use local executor for low-resource jobs	2 days ago
nextflow.config	Runspecs parameters to control core and node jobs	3 days ago

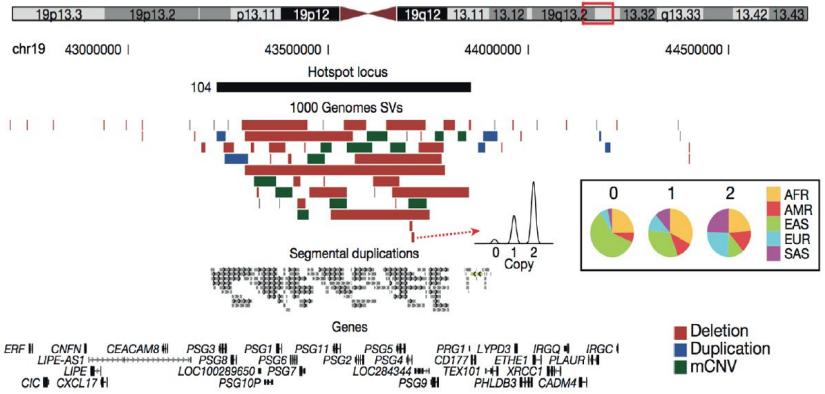
Downstream work

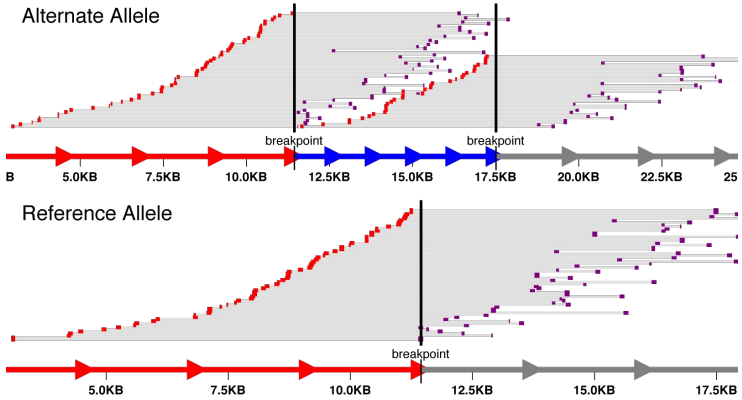
Genotyping SVs for cohorts

- Genome strip
- Clustering
- Remove benign events
- Genotyping based on function

Functional annotation and visualization

- SNPEff, vep, iAnnotateSV, svviz ...





http://svviz.readthedocs.io/en/latest/