Supplemental material to:
Optical genome mapping enables accurate testing of large repeat expansions

Bart van der Sanden[1], Kornelia Neveling[1], Syukri Shukor[2], Michael D. Gallagher[2], Joyce Lee[2], Stephanie L. Burke[2], Maartje Pennings[1], Ronald van Beek[1], Michiel Oorsprong[1], Ellen Kater-Baats[1], Eveline Kamping[1], Alide A. Tieleman[3], Nicol C. Voermans[3], Ingrid E. Scheffer[4], Jozef Gecz[5, 6, 7], Mark A. Corbett[7], Lisenka E.L.M. Vissers[1], Andy Wing Chun Pang[2], Alex Hastie[2], Erik-Jan Kamsteeg[1,*], Alexander Hoischen[1,8,*]

1 Department of Human Genetics, Research Institute for Medical Innovation, Radboud University Medical Center, Nijmegen, the Netherlands
2 Bionano Genomics Clinical and Scientific Affairs, San Diego, CA, USA
3 Department of Neurology, Radboud university medical center, Nijmegen, the Netherlands
4 Epilepsy Research Centre, Department of Medicine, University of Melbourne, Austin Health, VIC, Australia; Department of Pediatrics, University of Melbourne, Royal Children's Hospital, Florey and Murdoch Children's Research Institutes, VIC, Melbourne, Australia.
5 South Australian Health and Medical Research Institute, Adelaide, SA, Australia
6 Genetics and Molecular Pathology, SA Pathology, Adelaide, SA, Australia
7 Robinson Research Institute and Adelaide Medical School, University of Adelaide, Adelaide South Australia, Australia
8 Department of Internal Medicine, Radboud Expertise Center for Immunodeficiency and Autoinflammation and Radboud Center for Infectious Disease (RCI), Radboud university medical center, Nijmegen, the Netherlands
* These authors jointly supervised the work

Corresponding Authors
Erik-Jan Kamsteeg, PhD
Department of Human Genetics
Radboud university medical center
6500HB Nijmegen, The Netherlands
Email: erik-jan.kamsteeg@radboudumc.nl

AND

Alexander Hoischen, PhD
Department of Human Genetics & Department of Internal Medicine
Radboud university medical center
6500HB Nijmegen, The Netherlands
Email: alexander.hoischen@radboudumc.nl

TABLE OF CONTENTS

Supplemental Table S1: Deviation between the manual *de novo* assembly workflow and the local guided assembly workflow. For each data point in Fig 2, this table presents the estimated size of the largest allele resulting from both workflows, as well as the deviation between the two sizing workflows. Deviation is calculated as: ((local-manual)/local)*100. For this analysis the deviation was only calculated one way.

Sample	Manual de novo assembly	Local guided assembly	Deviation (%)[#]
CNBP_02	6331	6544	3.4
CNBP_03	6517	5374	-17.5
CNBP_04	7155	5652	-21.0
CNBP_05	8042	6322	-21.4
CNBP_07	5212	5173	-0.7
CNBP_08	2502	2661	6.3
CNBP_10	2874	3980	38.5
CNBP_11	375	254	-32.2
CNBP_12	3471	3346	-3.6
CNBP_13	4634	4330	-6.6
CNBP_14	5244	4331	-17.4
CNBP_15	2183	2092	-4.2
CNBP_16	3221	4747	47.4
CNBP_17	6275	5759	-8.2
CNBP_18	1915	2976	55.4
CNBP_19	1460	1574	7.8
CNBP_20	3977	4852	22.0
CNBP_21	288	244	-15.4
CNBP_22	1683	1725	2.5
CNBP_23	2131	2515	18.0
CNBP_25	1476	2626	77.9
CNBP_26	3737	3241	-13.3

Sample	Manual de novo assembly	Local guided assembly	Deviation (%)[#]
DMPK_01	269	247	-8.1
DMPK_02	456	473	3.8
DMPK_03	252	116	-54.0
DMPK_04	66	60	-9.1
DMPK_05	457	485	6.1
DMPK_06	64	82	28.1
DMPK_07	378	358	-5.2
DMPK_09	71	68	-4.2
DMPK_10	2829	2825	-0.1
DMPK_11	233	231	-1.0
DMPK_12	213	219	3.0
DMPK_13	163	167	2.7
DMPK_14	202	188	-6.9
DMPK_15	1839	1768	-3.9
DMPK_16	85	52	-38.6

DMPK_17	491	510	3.8
DMPK_18	71	69	-2.4
DMPK_19	54	61	13.7
DMPK_21	1366	1347	-1.4
DMPK_23	372	369	-0.9
DMPK_24	82	93	13.4
DMPK_25	79	109	38.0
DMPK_28	440	393	-10.6
DMPK_29	320	310	-3.1
DMPK_30	290	131	-54.8

Sample	**Manual de novo assembly**	**Local guided assembly**	**Deviation (%)[#]**
RFC1_01	1487	1497	0.6
RFC1_02	738	751	1.8
RFC1_03	883	897	1.6
RFC1_04	750	760	1.3
RFC1_05	1167	1175	0.7
RFC1_06	1565	1579	0.9
RFC1_07	625	643	2.9
RFC1_08	840	856	1.9
RFC1_09	1506	1499	-0.5
RFC1_10	1289	1307	1.4
RFC1_11	812	818	0.7
RFC1_12	1106	1121	1.4
RFC1_13	927	933	0.6
RFC1_14	725	740	2.0
RFC1_15	873	887	1.6
RFC1_16	1104	1097	-0.6
RFC1_17	711	723	1.6
RFC1_18	1444	1474	2.1
RFC1_19	223	235	5.6
RFC1_20	494	502	1.6
RFC1_21	703	715	1.6
RFC1_22	1161	1180	1.6
RFC1_23	1028	1054	2.5
RFC1_24	602	615	2.1
RFC1_25	855	868	1.6
RFC1_26	714	734	2.8
RFC1_27	973	987	1.5
RFC1_28	875	893	2.0
RFC1_29	1071	1073	0.2
RFC1_30	743	754	1.5

for calculating the average deviations in the manuscript, all deviations presented in this table were taken as positive value.

Supplemental Table S2: Repeat expansion sizes in repeat units at the three different repeat loci in ten control samples based on the manual *de novo* workflow. OGM only identified repeat sizes below the pathogenic repeat sizing threshold of *CNBP* and *DMPK*. In the case of *RFC1*, OGM identified three heterozygous and one homozygous expansion beyond the pathogenic repeat size threshold of 20 units that was used in this manuscript. However, the 20 units we used is different from the formal pathogenic repeat size threshold of *RFC1* at >400 units (**Methods** and **Table 3**). None of the calls exceeded this formal threshold and therefore the four samples are considered carriers. This data suggests that OGM has a very low false positive rate for the detection of repeat expansions.

	CNBP		***DMPK***		***RFC1***	
Control #	**Allele 1**	**Allele 2**	**Allele 1**	**Allele 2**	**Allele 1**	**Allele 2**
1	3	3	23	23	-3	-3
2	6	-39	30	30	59	13
3	17	17	14	14	-3	-4
4	25	-53	15	15	117	117
5	-37	-37	-17	-17	-1	-1
6	-26	-26	35	35	71	24
7	55	55	-11	-11	0	0
8	-5	-5	13	13	-4	-4
9	-42	-42	21	21	106	-1
10	-17	-17	-31	-46	14	14

Supplemental Table S3: The OGM results of the 12 additional samples with a known repeat expansion in *ATXN10*, *C9orf72*, *FXN*, *NOP56* or *STARD7.* For all samples except ATXN10_03, OGM detected a repeat expansion. In addition, OGM allowed to distinguish between the two repeat alleles for FXN_02. ATXN10 (Morato Torres et al. 2022) and C9orf72 (Barseghyan et al. 2022) repeat cases were also published before. For the molecule distance script "A" denotes multiple consensus maps and "B" denotes a gradient in the molecule distances. We now considered somatic instability in cases where both "A + B" provided suggestive evidence.

Sample	**Repeat motif**	**Pathogenic repeat size threshold**	**SOC**	**Manual *de novo* assembly**	**Local guided assembly**	**Suggestive evidence of somatic instability from molecule distance script**
ATXN10_01	ATTCT	>850	>850	-10 / 1,249	0 / 1,206	A
ATXN10_02	ATTCT	>850	>850	-11 / 1,076	0 / 1,056	A
ATXN10_03	ATTCT	>850	>850	-4 / 1,257	0 / 1,263	-
ATXN10_04	ATTCT	>850	>850	-10 / 1,249	2 / 2,711	A + B
C9orf72_01	GGCCCC	>250	>250	3 / 1,579	15 / 1,615	B
C9orf72_02	GGCCCC	>250	>250	7 / 2,314	16 / 16	A + B
FXN_01	GAA	>65	8 / >65	22 / 483	33 / 501	A
FXN_02	GAA	>65	>65 / >65	221 / 886	172 / 912	A
NOP56_01	GGCCTG	>670	5 / >40	16 / 1,081	19 / 1,077	A
NOP56_02	GGCCTG	>670	5 / >40	19 / 1,189	0 / 9	A + B
STARD7_01	ATTTC	>660	1,135*	-1 / 1,257	0/ 1,258	-

* SOC for *STARD7* was performed using long-range PCR followed by Oxford Nanopore sequencing.

Supplemental Table S4: Overview of which repeat expansions can be detected using our OGM analysis approach. The table contains all known repeat expansion loci published in (Depienne and Mandel 2021). "Yes" means the pathogenic repeat size threshold is >500 bp and repeats at this locus can be detected using our approach. "Possible" means the pathogenic repeat size threshold starts <300 bp, but extends beyond the 300 bp or even 500 bp mark. For these loci it depends on the actual size of the repeat expansion whether our approach works for these genes.

"No" means the pathogenic repeat size threshold remains <300 bp and repeat expansions at this locus cannot be detected using the OGM approach. For each gene the respective disorder, repeat unit, pathologic repeat size threshold is given. Table adjusted from (Depienne and Mandel 2021).

Gene	**Disorder**	**Repeat motif**	**Pathologic repeat size threshold (units)**	**Pathologic repeat size threshold (base pairs)**	**OGM capable to detect pathogenic repeat expansion**
AFF2	Fragile XE syndrome	CCG	≥200-900	≥600-2700	Yes
AR	Spinal and bulbar muscular atrophy	CAG	≥38-68	≥114-204	No
ARX	Early infantile epileptic encephalopathy type 1	GCN	23	69	No
ATN1	Dentatorubral-pallidoluysian atrophy	CAG	≥48-93	≥144-279	No
ATXN1	Spinocerebellar ataxia type 1	CAG	≥39-88	≥117-264	No
ATXN10	Spinocerebellar ataxia type 10	ATTCT/ATTGT	>280-4500	>1400-22500	Yes
ATXN2	Spinocerebellar ataxia type 2	CAG	≥32-500	≥96-1500	Possible
ATXN3	Spinocerebellar ataxia type 3	CAG	≥55-87	≥165-261	No
ATXN7	Spinocerebellar ataxia type 7	CAG	≥37-460	≥111-1380	Possible
ATXN8/ATXN8OS	Spinocerebellar ataxia type 8	CAG/CTG	>74-250	>222-750	Possible
BEAN1/TK2	Spinocerebellar ataxia type 31	TGGAA	≥110-760	≥550-3800	Yes
C9ORF72	Frontotemporal dementia / amyotrophic lateral sclerosis	GGGGCC	>30	>180	No
CACNA1A	Spinocerebellar ataxia type 6	CAG	≥20-33	≥60-99	No
CNBP	Myotonic dystrophy type 2	CCTG/CAGG	>50-11000	>200-44000	Possible
CSTB	Progressive myoclonus epilepsy type 1	CCCCGCCCCGCG	30-75	360-900	Possible
DAB1	Spinocerebellar ataxia type 37	ATTTC	≥31-75	≥155-375	Possible
DMPK	Myotonic dystrophy type 1	CTG	>50-10000	>150-30000	Possible
FMR1	Fragile X syndrome	CGG	>200	>600	Yes
FMR1	Fragile X-associated premature ovarian infertility	CGG	55-200	165-600	Possible

FMR1	Fragile X-associated tremor ataxia syndrome	CGG	55-200	165-600	Possible
FOXL2	Blepharophimosis, ptosis and epicanthus inversus	GCN	19-24	57-72	No
FXN	Friedreich ataxia	GAA	≥66-1300	≥198-3900	Possible
GIPC1	Oculopharyngeal muscular dystrophy type 2	CGG	≥97-120	≥291-360	Possible
GLS	Global developmental delay, progressive ataxia, and elevated glutamine	GCA	≥680-1400	≥2040-4200	Yes
HOXA13	Hand-foot-genital syndrome	GCN	22	66	No
HOXD13	Synpolydactyly type 1	GCG	24	72	No
HTT	Huntington disease	CAG	≥36-250	≥108-750	Possible
JPH3	Huntington disease-like 2	CAG	≥41-58	≥123-174	No
LRP12	Oculopharyngodistal myopathy type 1	CGG	90-130	270-390	Possible
MARCHF6	Familial adult myoclonic epilepsy type 3	ATTTC	≥660-2800	≥3300-14000	Yes
NOP56	Spinocerebellar ataxia type 36	GGCCG	≥650-2500	≥3250-12500	Yes
NOTCH2NLC	Neuronal intranuclear inclusion disease	CGG	≥61-500	≥183-1500	Possible
NUM2B-AS1	Oculopharyngeal myopathy with leukoencephalopathy type 1	CGG/CCG	40-60	120-180	No
PABPN1	Oculopharyngeal muscular dystrophy	GCG	≥12-17	≥36-51	No
PHOX2B	Congenital central hypoventilation syndrome	GCN	25-29	75-87	No
PPP2R2B	Spinocerebellar ataxia type 12	CAG	43-78	129-234	No
RAPGEF2	Familial adult myoclonic epilepsy type 7	ATTTC	N/A	N/A	N/A
RFC1	Cerebellar ataxia, neuropathy and vestibular areflexia syndrome	AAGGG	≥400-2000	≥2000-10000	Yes
RUNX2	Brachydactyly and cleidocranial dysplasia	GCN	27	81	No
SAMD12	Familial adult myoclonic epilepsy type 1	ATTTC	≥440-3680	≥2200-18400	Yes
SOX3	Mental retardation with isolated growth hormone deficiency	GCN	26	78	No

STARD7	Familial adult myoclonic epilepsy type 2	ATTTC	≥661-735	≥3305-3675	Yes
TAF1	X-linked dystonia parkinsonism	CCCTCT	30-55	180-330	Possible
TBP	Spinocerebellar ataxia type 17	CAG or CAG/CAA	≥43-66	≥129-178	No
TCF4	Fuchs endothelial corneal dystrophy type 3	CTG	>50	>150	No
TNRC6A	Familial adult myoclonic epilepsy type 6	ATTTC	N/A	N/A	N/A
XYLT1	Baratela-Scott syndrome	CGG	120-800	360-2400	Yes
YEATS2	Familial adult myoclonic epilepsy type 4	ATTTC	N/A	N/A	N/A
ZIC2	Holoprosencephaly type 5	GCN	25	75	No

Supplemental Table S5: Genomic regions and label IDs used for the manual *de novo* assembly size estimate. Table adjusted from van der Sanden et al. (2024).

Gene	Genomic region of repeat expansion locus (GRCh38/hg38)	Start label of interest ID[#]	End label of interest ID[#]	Reference length between labels of interest (bp)
CNBP	chr3:129,169,450-129,181,839	26,243	26,246	12,389
DMPK	chr19:45,752,584-45,771,947	5,926	5,927	19,363
RFC1	chr4:39,343,732-39,350,590	7,723	7,724	6,858

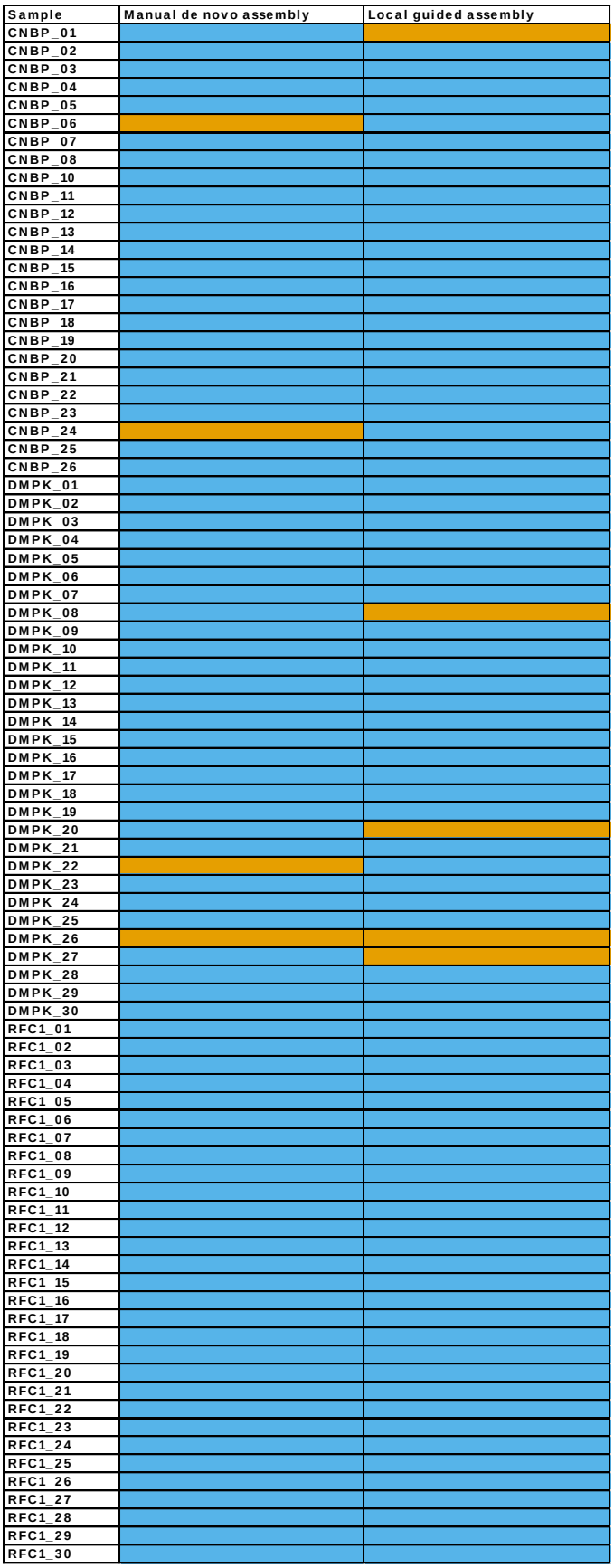
The label IDs are chromosome specific.
For *CNBP* the locus entailed four labels to prevent resolution errors from labels in proximity that can be falsely read as one extra-bright signal.

Supplemental Table S6: Genomic regions and label IDs used for the Molecule Distance Script workflow. Table adjusted from van der Sanden et al. (2024).

Gene	**Genomic region of repeat expansion locus (GRCh38/hg38)**	**Start label of interest ID#**	**End labelof interest ID#**	**Reference length between labels of interest (bp)**
CNBP	chr3:129,168,220-129,186,501	26,242	26,247	18,281
DMPK	chr19:45,740,976-45,829,452	5,925	5,930	88,476
RFC1	chr4:39,339,156-39,362,887	7,722	7,725	23,731

The label IDs are chromosome specific.

For all three genes the locus entailed more than two labels to make sure the labels of interest cover the entire gene region containing the repeat expansion since the molecule distance script workflow constitutes an automated workflow and does not allow for individual manual tweaking.



Supplemental Fig S1: Heatmap presenting the individual results from each OGM workflow. Blue color means repeat expansion detected and orange color means no repeat expansion detected.

Supplemental References

Barseghyan H, Pang AWC, Zhang Y, Sahajpal NS, Delpu Y, Lai C-YJ, Lee J, Tessereau C, Oldakowski M, Kolhe RB et al. 2022. Neurogenetic Variant Analysis by Optical Genome Mapping for Structural Variation Detection-Balanced Genomic Rearrangements, Copy Number Variants, and Repeat Expansions/Contractions. In *Genomic Structural Variants in Nervous System Disorders*, doi:10.1007/978-1-0716-2357-2_9 (ed. C Proukakis), pp. 155-172. Springer US, New York, NY.

Depienne C, Mandel J-L. 2021. 30 years of repeat expansion disorders: What have we learned and what are the remaining challenges? *The American Journal of Human Genetics* **108**: 764-785.

Morato Torres CA, Zafar F, Tsai YC, Vazquez JP, Gallagher MD, McLaughlin I, Hong K, Lai J, Lee J, Chirino-Perez A et al. 2022. ATTCT and ATTCC repeat expansions in the ATXN10 gene affect disease penetrance of spinocerebellar ataxia type 10. *HGG Adv* **3**: 100137.

van der Sanden B, Neveling K, Pang AWC, Shukor S, Gallagher MD, Burke SL, Kamsteeg E-J, Hastie A, Hoischen A. 2024. Optical Genome Mapping for Applications in Repeat Expansion Disorders. *Current Protocols* **4**: e1094.