Published in final edited form as:
Curr Protoc Hum Genet. ; 80: 7.22.1–7.22.9. doi:10.1002/0471142905.hg0722s80.

Design of Large-Insert Jumping Libraries for Structural Variant Detection using Illumina Sequencing

C. Hanscom[1] and **M. Talkowski**[1,2]

[1]Molecular Neurogenetics Unit, Psychiatric and Neurodevelopmental Genetics Unit, Center for Human Genetic Research, Massachusetts General Hospital, Boston, Massachusetts

[2]Department of Neurology, Harvard Medical School, Boston, Massachusetts

Abstract

Next generation sequencing is an important and efficient tool for the identification of structural variation, particularly balanced chromosomal rearrangements because such events are not routinely detected by microarray and localization of altered regions by karyotype is imprecise. Indeed, the degree of resolution that can be obtained through next generation technologies enables elucidation of precise breakpoints and has facilitated the discovery of numerous pathogenic loci in human disease and congenital anomalies. The protocol described here explains one type of large-insert ‘jumping library’ and the steps required to generate them for multiplexed sequencing using Illumina sequencing technology. This approach allows for costefficient multiplexing of samples and derives a very high yield of fragments with large insets, or ‘jumping’ fragments.

Keywords

next generation sequencing; mate pair sequencing; jumping libraries; structural variation

INTRODUCTION

There have been several different approaches described to perform massively parallel sequencing of large fragments on multiple platforms. These methods have been referred to under various terms, such as large-insert sequencing, mate-pair sequencing, and jumping libraries, which all represent similar approaches to sequencing large genomic inserts. In the protocol described below, we provide one application of these methods, which is an adaptation of the mate-pair library protocol used with the Applied Biosystems SOLiD sequencing platform (Applied Biosystems SOLiD™ 4 System) but adapted for Illumina sequencing. We chose this protocol because in our experience it is the most efficient in terms of lowest cost and highest proportion of large inserts generated, though there is a tradeoff in terms of the length of the sequence fragments generated. Multiplexing is accomplished by using individual Y shaped adapters, which contain Illumina compatible barcoding sequence on one strand. This protocol and several variations of this process have

To whom correspondence should be addressed Michael E. Talkowski, Ph.D., Center for Human Genetic Research, Massachusetts General Hospital and Harvard Medical School, 185 Cambridge St, Boston, MA 02114, talkowski@chgr.mgh.harvard.edu.

used previously for elucidation of structural variation (Korbel et al., 2007; Talkowski et al., 2011; Talkowski and Rosenfeld et al., 2012; Chiang et al., 2012).

BASIC PROTOCOL 1

Generation of Libraries with Large Genomic Inserts for 25BP Paired End Illumina Platform Sequencing

The purpose of this method is to describe a protocol for creating DNA libraries suitable for 25 cycles of next generation sequencing. The method derives short fragments appropriate for massively parallel sequencing on an Illumina platform (Illumina Inc) that are separated by large genomic inserts of a user-selected size. As detailed above, a variation of the jumping library method was first applied in next-generation sequencing by Snyder and colleagues in their seminal structural variation sequencing paper using 454 sequencing technology, and similar methods have been released by Illumina in which circularization is achieved by blunt end ligation (Korbel et al., 2007; Bentley et al., 2008). Applied Biosystems developed a derivation of this protocol using an internal adaptor for circularization and restriction digestion to fragment the circle, followed by sample preparation for ABI SOLiD sequencing (web resources) and we have previously described modifications to that protocol for Illumina sequencing by synthesis technology (Talkowski et al., 2011), which we enumerate in detail here. In overview, genomic DNA is sheared to a targeted insert size, circularized, then fragmented with the circularization junctions retained, thereby deriving short genomic fragments in which the ends are separated by the size of the circle. This method enables massively parallel sequencing of the ends of the fragments and leveraging the size of the insert into information about the assembly of the genome. We and others have used these methods to identify structural variation in the genome, including the complex reorganization of ‘shattered’ chromosomal segments that has been defined as chromothripsis, to delineate transgenic integration sites, to identify genes disrupted by balanced structural variation that contribute to human disease (Chen et al., 2010; Kloosterman et al., 2011; Talkowski et al., 2011; Talkowski and Maussion et al., 2012; Chiang et al., 2012; Hodge et al., 2013). We have also shown the feasibility of this method to characterize structural variation in prenatal diagnostic testing (Talkowski and Ordulu et al., 2012).

Materials

Reagents	Vendor/cat.no.
5-10ug genomic DNA sample	sample
1X TE Buffer	12090-015
QIAquick PCR Purification Kit	28104 or 28106
End-It DNA End-Repair Kit	Epicentre, ER81050
RNase-Free Duplex Buffer	IDT
Cap adapter oligo duplex	see recipe
Quick ligation kit	NEB, M2200S or M2200L
UltraPure Agarose	Life Technologies, 16500-500
Ethidium Bromide	Sigma, E1510-10ML
1 kb+ DNA ladder	Life Technologies, 10787-018
QIAquick Gel Extraction Kit	28704 or 28706

Quant-iT PicoGreen dsDNA Assay Kit	Life Technologies, P7589
internal circularization adapter duplex	see recipe
Plasmid-safe DNASE kit	Epicentre, E3105K
EcoP15I restriction enzyme	NEB, R0646L
BSA 100X	NEB, B9000S
10mM ATP (contained in NEB EcoP15I package)	NEB
NEB Buffer 3 (contained in NEB EcoP15I package)	NEB
Sinefungin, 10mM	Millipore, 567051-2MG
Klenow DNA polymerase I (lg) fragment	NEB, M0210S
Nuclease-free water	Ambion, AM9930
Sodium Chloride(5M)	Boston BioProducts, BM-244
Tris-HCL Buffer (1M, pH 7.5)	Boston BioProducts, BM-315
EDTA (0.5M, pH 8.0)	Boston BioProducts, BM-150
Dynabeads MyOne Streptavidin C1	65001
NEBNext dA Tailing Module	NEB, E6053L
barcoded Y adapter duplex	see recipe
Phusion High-Fidelity PCR MM w/HF Buffer	NEB, M0531S
Agilent 1000 kit	5067-4626

Equipment	**Vendor/cat.no.**
NanoDrop spectrophotometer	ND1000
Covaris Focused-ultrasonicator (S-Series or E-Series)	Covaris
Covaris miniTUBES (blue) for 3kb shearing	520065
Centrifuge	Eppendorf 5804
Gel electrophoresis apparatus	BIO-RAD
Vacuum Manifold (optional)	Qiagen
ThermoCycler	BIO-RAD C1000
Heat Block or incubator (37C, 65C)	eppendorf
Magnetic Rack	Invitrogen CS15000
Agilent 2100 Bioanalyzer	Agilent 2100 or equivalent QC method

Fragmentation of Human Genomic DNA

1. Use the Covaris S2 instrument to shear approximately 5-10ug of high quality human genomic DNA to a target size of 3kb using Covaris' nominal shear protocol.
2. Follow the manufacturer's instructions for the Qiagen PCR Purification Kit (or, where applicable, Qiagen Gel Purification) for each reaction cleanup. The elution volume for this step is 35ul per 5ug starting material.

   It is important that the initial DNA sample is of high-quality and, ideally, at least 5ug in quantity. Poor quality, degraded DNA and/or insufficient starting material can result in both poor library yield and high duplication rates. The following steps are optimized for 10ug of starting material. Reaction volumes should be scaled proportionally for an initial quantity of 5ug.

End Repair of Sheared DNA

3 Use the Epicentre's EndIt End Repair kit (or equivalent blunt ending method suitable for subsequent ligation) to end repair the sheared DNA fragments. One reaction volume (50ul) is used per 5ug starting material. Qiagen column purification, 50ul elution volume.

EcoP15I Cap Adapter Ligation

4 Determine the concentration of sheared, end-repaired DNA using a Nanodrop instrument (see Appendix 3D). Determine the volume of duplexed 50uM cap adapter required for an adapter: fragment ratio of 10:1.

Note: Initial preparation of the capping oligos, as well as all other double stranded adapters used within this protocol, is achieved by resuspending both strands at the desired concentration in duplexing buffer (IDT RNase-Free Duplex Buffer). Oligos are briefly denatured by heating to 94C for 2minutes, and then allowed to cool to form double-stranded adapters. Once prepared, duplexes may be stored at -20C until needed.

5 Use the NEB Quick Ligation Kit reagents and protocol to ligate the cap adapters.

Gel Size Selection

6 Prepare a 0.8-1% agarose (EtBr, 2.5ul per 150ml) gel to use for more specific size selection.

7 Add 10ul 6X loading dye to the sample and load the entire product across two lanes. If 5ug of starting material was used, a proportionate volume can be loaded into a single lane. Run the gel until a bromophenol dye indicator has migrated approximately 2-3cm.

Running the gel longer than this may result in the need to use more than one qiaquick column due to agarose mass.

8 Extract the product within the range of 2.5-6kb on the gel and purify with Qiagen reagents, 50ul elution volume.

Circularization

9 Use NanoDrop or PicoGreen measurements to determine the volume of duplexed 2uM internal adapter required for an adapter: fragment ratio of 3:1.

An ideal yield at this stage is around 1ug of DNA (per 5ug starting material), but libraries can successfully be prepared with 0.5ug or less.

If NanoDrop measurements are used, a strong solvent absorption peak <230nm will usually be present. Comparison with PicoGreen measurements at has found that readings remain adequately reliable for this step. PicoGreen measurement is recommended if the solvent absorption maximum occurs at a wavelength greater than 230nm.

10 The previously ligated cap adapters contain AC overhangs which are complimentary to the overhangs of the internal adapter. The gel purified product is circularized upon this adapter, which contains a biotinylated thymine that will be later used for retention of the circularization junction. Illumina recommends a dilute DNA concentration (2ng/ul) in order to help reduce the number of chimeric circularizations. Use NEB quick ligation reagents to set up the following 500ul reaction per 1ug DNA (scale reaction volumes as needed):

DNA(1ug)	50ul
NEB quick ligase buffer:	250ul
H_2O:	187ul
Internal (2uM)	0.76ul
NEB ligase:	12ul

Qiagen column purification, 60ul elution volume.

DNase Treatment

11 Treat the circularized product with DNase using Epicentre's Plasmid Safe Dnase reagents to remove non-circularized products. Qiagen column purification, 30ul elution volume.

EcoP15I Digest

12 Use the following reaction setup to digest overnight (min 12 hour) at 37C:

EcoP15I Digest (100ul reaction)

Circularized DNA	30ul
10X NEB Buffer3	10ul
10X BSA	10ul
10X ATP(25nM)	20ul
H2O	23ul
Sinefungin (10mM)	1ul
EcoP15I(10U/ul)	6ul

Following the 12 hour incubation, add 2uL 25mM ATP (NEB) and 0.5 uL EcoP15I enzyme and incubate for an additional hour, then incubate at 65C for 20 minutes to heat inactivate enzymes.

End Repair of Digested DNA

13 Add 1.5uL 25mM (each) dNTP mix and 1ul Klenow (large) fragment to the reaction mixture to end repair the digested DNA. Allow the end repair reaction to proceed for 40min at room temperature, then heat inactivate again by incubating at 65C for another 20 minutes. Cool reactions on ice for 5 minutes.

Streptavidin Bead Binding

14 See recipe for preparation of wash and binding buffers. Wash 35ul of MyOne C1 beads for each library preparation with bead wash buffer. As long as volume

permits, the total required volume of beads can be washed in a single tube. Add 500ul wash buffer to the beads, vortex the tube briefly to homogenize the mixture, and then quick centrifuge to remove any liquid from the tube cap. Use a magnetic rack to retain the beads while removing supernatant. Repeat for a total of three washes.

15 Resuspend the washed beads in their original volume of 1X binding buffer (1 part binding buffer as per recipe, 1 part nuclease free water).

16 Add 105ul binding buffer and 35ul washed bead solution to each ~105ul DNA sample. Bind for approx 30 minutes at room temperature with periodic mixing.

17 After binding, wash again three times with 500ul wash buffer.

It is important to remove as much wash buffer as possible in this and all other final wash steps, as residual wash solution can interfere with subsequent enzymatic reactions. It is helpful to quickly spin the tubes after the final wash to draw remaining liquid to the bottom for easier removal of residual wash solution.

dA Tailing of DNA (on beads)

18 Use NEB dA tailing module to add a single A base to the end repaired DNA, following the manufacturer's protocol for 50ul reactions. After the reactions are complete, wash beads again three times with 500ul wash buffer. A final quick centrifugation is helpful in removing remnants of wash buffer.

Adapter Ligation (on beads)

19 Use NEB Quick Ligation Kit to ligate custom barcode adapters to the DNA fragments. Use 0.34uL 15uM duplexed adapter per 1ug circularized DNA (as measured in step9). After the reaction is complete, wash beads again three times with 500ul wash buffer, and once with 500ul EB. A final quick centrifugation is helpful in removing remnants of buffer. Resuspend cleaned reactions in 30ul EB.

PCR Amplification (on beads)

20 Three PCR reactions are run per library using the setup and protocol indicated in Support Protocol 1.

Gel Purification of PCR products

21 Use a magnetic stand to separate post-PCR solution from the beads and purify on one Qiagen PCR purification column, 30ul elution volume. Load the entire elution volume on a single lane of 2% agarose (EtBr 2.5ul/150ml gel) gel. Some variation in yield between libraries amplified with the same number of cycles is normal, but product should be clearly visible as a distinct ~200BP band. Extract this product using Qiagen gel purification reagents. Elute with 17ul elution volume and add 1.5ul 1% TWEEN20.

If peaks lower than ~200BP are present, run the gel longer in order to avoid possible contamination with dimer during gel extraction.

QC of Final Product

22 Assess gel purified products using QPCR, Agilent Bioanalyzer 2100 or Agilent TapeStation 2200 methods.

SUPPORT PROTOCOL 1: PCR protocol for library amplification

After barcode adapters are ligated and reactions washed, libraries are amplified by PCR using a universal forward primer and a reverse primer dependent upon the specific barcode used. Three PCR reactions are performed per sample in order to reduce random amplification bias. An initial cleanup prior to final gel purification concentrates the sample for gel loading and also removes excess primers and dNTPs from solution.

Materials

Barcode-ligated DNA sample (on beads, resuspended in 30ul EB)

Phusion High-Fidelity PCR MM w/HF Buffer (NEB, M0531S)

Univ_PCR forward primer (25uM)

Index compatible PCR reverse primer (25uM)

Nuclease free H2O

1. For each library, prepare three 50ul PCR reactions using the reagents and ratios indicated. PCR reagents for three 50ul reactions:

DNA (on beads)	30ul
Primer 1.0 (25uM)	1ul
Index Primer (25uM)	1ul
2X Phusion mix	75ul
Nuclease free water	43uL

2. Amplify libraries using the indicated PCR protocol.

98C	30 seconds
98C	10 seconds
65C	30 seconds
72C	30 seconds X 8-12 total cycles
72C	5 min, 10C hold

3. Following amplification, use a magnetic rack to remove the supernatant from the beads. Pool each set of three reactions into a single cleanup.

REAGENTS AND SOLUTIONS

Streptavidin Binding and Wash Buffers (scale volumes as needed):

Tris-HCL 1M PH-7.5	60ul
NaCl 5M	2.4ml

EDTA 0.5M	12ul
for wash buffer:	bring volume to 12ml with nuclease free water and add 1ul TWEEN20.
for binding buffer:	bring volume to 6ml with nuclease free water.

Buffers can be stored at room temperature for three months.

Adapter Information (5-3'):

EcoP15I cap adapter / initial circularization adapter (duplex, 50uM):

Cap_adapter_1:	/5Phos/ACAGCAG
Cap_adapter_2:	/5Phos/CTGCTGTAC

Circularization Adapter (deplex, 2uM):

Internal_1A:	/5Phos/CGT TC/iBiodT/CCG T
Internal_2A:	/5Phos/GGA GAA CGG T

Barcoded Adapters (duplex, 15uM) :

UniversalMPX:	ACACTCTTTCCCTACACGACGCTCTTCCGATC*T
MPX_index##:	/5Phos/GATCGGAAGAGCACACGTCTGAACTCCAGTCAC(+6BPindex)

*Phosphorothioate (S-oligos) are used to help prevent degradation

PCR Primers (25uM):

Univ_PCR:	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC*T
Rev_index#:	CAAGCAGAAGACGGCATACGAGAT(+6BPprimerindex**)GTGACTGGAGTTC

**rev primer index is the reverse compliment of the barcode adapter index

Commentary

Background

The purpose of this protocol is to describe the preparation of DNA libraries suitable for 25 cycles of next generation sequencing. This method, which is compatible with the Illumina platform (Illumina Inc), results in libraries consisting of short DNA fragments that represent junctions formed from the circularization of larger (~3kb) genomic fragments. The use of short fragments derived from long genomics inserts allows higher effective genomic coverage while minimizing the cost of whole-genome sequence coverage, and this cost savings will scale with future decreases in sequencing costs.

Critical Parameters

It is important to assess both the quantity and quality of the starting material, as these factors are the two that most significantly influence the quality of finished libraries. Visualizing a genomic sample (~100-200ng) on a 1% agarose gel to determine DNA quality is recommended, as degraded DNA will result in both poor library yield and high duplication rates. Genomic DNA should visualize as a distinct high molecular weight band. There should be little to no visible smearing present. PicoGreen, or another method more specific

to doublestranded DNA, is recommended for initial quantification because 260nm absorption is an imprecise method for measuring the concentration of intact, dsDNA. In addition to initial DNA quantity, other factors that influence the quality of a library are reaction efficiency, number of PCR cycles, and the final gel purification. Residual wash buffer, used in steps 14-16, can interfere with the enzymatic reactions associated with these steps. It is important that residual wash buffer be removed for maximum reaction efficiency.

The number of PCR cycles in the final amplification influences the degree of amplification bias. Using fewer cycles will result in libraries with less bias, however too few will not provide DNA in sufficient quantity. DNA yield following the initial gel size selection provides a good way to initially estimate the number of cycles to use: if concentration at this stage is roughly 1ug or higher, then 8-10 cycles should provide sufficient amplification.

The final purification steps are also important for removing traces of primer and/or primer dimer, which can interfere with sequencing, from the final libraries. The final gel should be run long enough to cleanly isolate the ~200BP band during gel extraction from dimer bands that (if present) run at ~130BP. The final products should run as a clearly visible, distinct ~200BP bands on an agarose gel and visualize as sharp peaks on Bioanalyzer or TapeStation traces. Final expected concentration can vary depending on starting quantity, reaction efficiencies, and PCR cycles used, but is usually between 5-100nM.

Time Considerations

Total protocol time can vary depending upon the number of samples being prepared and equipment available. For example, shearing time using Covaris 3kb nominal protocol is 10 minutes per sample. The time required for size selection and final gel purification also varies, since the number of samples that can be run on a single gel is limited. The amount of time needed to prepare, purify and wash reaction samples will also influence the overall schedule. The EcoP15I digest is completed overnight, so it is practical to time other preceding reactions around this step. In general, a small number of samples can be prepared in two days (day 1: steps 1-12; day 2: steps 13-22), whereas three or more days is a more realistic schedule for a larger number of samples. Eight samples are a convenient number to prepare at once, in which case this protocol can usually be completed in three days (day1: steps 1-12; day 2: steps 13-19; day 3: steps 20-23).

Anticipated Results

This protocol should deliver Illumina adapter ligated, small DNA fragments in which the ends are separated by the initial user-selected fragment size (~3 kb described here). Paired-end sequencing should yield predictable ‘jumps’ in the genome following alignment for a high proportion (>99% in our previous studies) of fragments sequenced, enabling routine delineation of structural variation or genome assembly.

Acknowledgments

We thank Shangtao Liu for helping to develop and implement this protocol. This work was funded by the National Institutes of Health (MH095867, HD065286, and GM061354).

INTERNET RESOURCES

http://tools.invitrogen.com/content/sfs/manuals/SOLiD4_Library_Preparation_man.pdf
Applied Biosystems SOLiD™ 4 System Library Preparation Guide, 2010. Web.

LITERATURE CITED

Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, Hall KP, Evers DJ, Barnes CL, Bignell HR, Boutell JM, Bryant J, Carter RJ, Keira Cheetham R, Cox AJ, Ellis DJ, Flatbush MR, Gormley NA, Humphray SJ, Irving LJ, Karbelashvili MS, Kirk SM, Li H, Liu X, Maisinger KS, Murray LJ, Obradovic B, Ost T, Parkinson ML, Pratt MR, Rasolonjatovo IM, Reed MT, Rigatti R, Rodighiero C, Ross MT, Sabot A, Sankar SV, Scally A, Schroth GP, Smith ME, Smith VP, Spiridou A, Torrance PE, Tzonev SS, Vermaas EH, Walter K, Wu X, Zhang L, Alam MD, Anastasi C, Aniebo IC, Bailey DM, Bancarz IR, Banerjee S, Barbour SG, Baybayan PA, Benoit VA, Benson KF, Bevis C, Black PJ, Boodhun A, Brennan JS, Bridgham JA, Brown RC, Brown AA, Buermann DH, Bundu AA, Burrows JC, Carter NP, Castillo N, Chiara E, Catenazzi M, Chang S, Neil Cooley R, Crake NR, Dada OO, Diakoumakos KD, Dominguez-Fernandez B, Earnshaw DJ, Egbujor UC, Elmore DW, Etchin SS, Ewan MR, Fedurco M, Fraser LJ, Fuentes Fajardo KV, Scott Furey W, George D, Gietzen KJ, Goddard CP, Golda GS, Granieri PA, Green DE, Gustafson DL, Hansen NF, Harnish K, Haudenschild CD, Heyer NI, Hims MM, Ho JT, Horgan AM, Hoschler K, Hurwitz S, Ivanov DV, Johnson MQ, James T, Huw Jones TA, Kang GD, Kerelska TH, Kersey AD, Khrebtukova I, Kindwall AP, Kingsbury Z, Kokko-Gonzales PI, Kumar A, Laurent MA, Lawley CT, Lee SE, Lee X, Liao AK, Loch JA, Lok M, Luo S, Mammen RM, Martin JW, McCauley PG, McNitt P, Mehta P, Moon KW, Mullens JW, Newington T, Ning Z, Ling Ng B, Novo SM, O'Neill MJ, Osborne MA, Osnowski A, Ostadan O, Paraschos LL, Pickering L, Pike AC, Pike AC, Chris Pinkard D, Pliskin DP, Podhasky J, Quijano VJ, Raczy C, Rae VH, Rawlings SR, Chiva Rodriguez A, Roe PM, Rogers J, Rogert Bacigalupo MC, Romanov N, Romieu A, Roth RK, Rourke NJ, Ruediger ST, Rusman E, Sanches-Kuiper RM, Schenker MR, Seoane JM, Shaw RJ, Shiver MK, Short SW, Sizto NL, Sluis JP, Smith MA, Ernest Sohna Sohna J, Spence EJ, Stevens K, Sutton N, Szajkowski L, Tregidgo CL, Turcatti G, Vandevondele S, Verhovsky Y, Virk SM, Wakelin S, Walcott GC, Wang J, Worsley GJ, Yan J, Yau L, Zuerlein M, Rogers J, Mullikin JC, Hurles ME, McCooke NJ, West JS, Oaks FL, Lundberg PL, Klenerman D, Durbin R, Smith AJ. Accurate whole human genome sequencing using reversible terminator chemistry. Nature. 2008; 456(7218):53–59. [PubMed: 18987734]

Chiang C, Jacobsen JC, Ernst C, Hanscom C, Heilbut A, Blumenthal I, Mills RE, Kirby A, Lindgren AM, Rudiger SR, McLaughlan CJ, Bawden CS, Reid SJ, Faull RL, Snell RG, Hall IM, Shen Y, Ohsumi TK, Borowsky ML, Daly MJ, Lee C, Morton CC, MacDonald ME, Gusella JF, Talkowski ME. Complex reorganization and predominant non-homologous repair following chromosomal breakage in karyotypically balanced germline rearrangements and transgenic integration. Nat Genet. 2012; 44(4):390–397. S391. [PubMed: 22388000]

Chen W, Ullmann R, Langnick C, Menzel C, Wotschofsky Z, Hu H, Döring A, Hu Y, Kang H, Tzschach A, Hoeltzenbein M, Neitzel H, Markus S, Wiedersberg E, Kistner G, van Ravenswaaij-Arts CM, Kleefstra T, Kalscheuer VM, Ropers HH. Breakpoint analysis of balanced chromosome rearrangements by next-generation paired-end sequencing. Eur J Hum Genet. 2010; 18(5):539–543. [PubMed: 19953122]

Hodge JC, Mitchell E, Pillalamarri V, Toler TL, Bartel F, Kearney HM, Zou YS, Tan WH, Hanscom C, Kirmani S, Hanson RR, Skinner SA, Rogers RC, Everman DB, Boyd E, Tapp C, Mullegama SV, Keelean-Fuller D, Powell CM, Elsea SH, Morton CC, Gusella JF, Dupont B, Chaubey A, Lin AE, Talkowski ME. Disruption of MBD5 contributes to a spectrum of psychopathology and neurodevelopmental abnormalities. Mol Psychiatry. 2013

Kloosterman WP, Guryev V, van Roosmalen M, Duran KJ, de Bruijn E, Bakker SC, Letteboer T, van Nesselrooij B, Hochstenbach R, Poot M, Cuppen E. Chromothripsis as a mechanism driving complex de novo structural rearrangements in the germline. Hum Mol Genet. 2011; 20(10):1916–1924. [PubMed: 21349919]

Korbel JO, Urban AE, Affourtit JP, Godwin B, Grubert F, Simons JF, Kim PM, Palejev D, Carriero NJ, Du L, Taillon BE, Chen Z, Tanzer A, Saunders AC, Chi J, Yang F, Carter NP, Hurles ME, Weissman SM, Harkins TT, Gerstein MB, Egholm M, Snyder M. Paired-end mapping reveals extensive structural variation in the human genome. Science. 2007; 318(5849):420–426. [PubMed: 17901297]

Talkowski ME, Ernst C, Heilbut A, Chiang C, Hanscom C, Lindgren A, Kirby A, Liu S, Muddukrishna B, Ohsumi TK, Shen Y, Borowsky M, Daly MJ, Morton CC, Gusella JF. Next-generation sequencing strategies enable routine detection of balanced chromosome rearrangements for clinical diagnostics and genetic research. Am J Hum Genet. 2011; 88(4):469–481. [PubMed: 21473983]

Talkowski ME, Maussion G, Crapper L, Rosenfeld JA, Blumenthal I, Hanscom C, Chiang C, Lindgren A, Pereira S, Ruderfer D, Diallo AB, Lopez JP, Turecki G, Chen ES, Gigek C, Harris DJ, Lip V, An Y, Biagioli M, Macdonald ME, Lin M, Haggarty SJ, Sklar P, Purcell S, Kellis M, Schwartz S, Shaffer LG, Natowicz MR, Shen Y, Morton CC, Gusella JF, Ernst C. Disruption of a large intergenic noncoding RNA in subjects with neurodevelopmental disabilities. Am J Hum Genet. 2012; 91(6):1128–1134. [PubMed: 23217328]

Talkowski ME, Ordulu Z, Pillalamarri V, Benson CB, Blumenthal I, Connolly S, Hanscom C, Hussain N, Pereira S, Picker J, Rosenfeld JA, Shaffer LG, Wilkins-Haug LE, Gusella JF, Morton CC. Clinical diagnosis by whole-genome sequencing of a prenatal sample. N Engl J Med. 2012; 367(23):2226–2232. [PubMed: 23215558]

Talkowski ME, Rosenfeld JA, Blumenthal I, Pillalamarri V, Chiang C, Heilbut A, Ernst C, Hanscom C, Rossin E, Lindgren AM, Pereira S, Ruderfer D, Kirby A, Ripke S, Harris DJ, Lee JH, Ha K, Kim HG, Solomon BD, Gropman AL, Lucente D, Sims K, Ohsumi TK, Borowsky ML, Loranger S, Quade B, Lage K, Miles J, Wu BL, Shen Y, Neale B, Shaffer LG, Daly MJ, Morton CC, Gusella JF. Sequencing chromosomal abnormalities reveals neurodevelopmental loci that confer risk across diagnostic boundaries. Cell. 2012; 149(3):525–537. [PubMed: 22521361]