



Molecular genetic decoding of malformations of cortical development

Jae Seok Lim and Jeong Ho Lee*

Graduate School of Medical Science and Engineering, KAIST, Daejeon, Korea

Malformations of cortical development (MCD) cover a broad spectrum of developmental disorders which cause the various clinical manifestations including epilepsy, developmental delay, and intellectual disability. MCD have been clinically classified based on the disruption of developmental processes such as proliferation, migration, and organization. Molecular genetic studies of MCD have improved our understanding of these disorders at a molecular level beyond the clinical classification. These recent advances are resulted from the development of massive parallel sequencing technology, also known as next-generation sequencing (NGS), which has allowed researchers to uncover novel molecular genetic pathways associated with inherited or *de novo* mutations. Although an increasing number of disease-related genes or genetic variations have been identified, genotype-phenotype correlation is hampered when the biological or pathological functions of identified genetic variations are not fully understood. To elucidate the causality of genetic variations, *in vivo* disease models that reflect these variations are required. In the current review, we review the use of NGS technology to identify genes involved in MCD, and discuss how the functions of these identified genes can be validated through *in vivo* disease modeling.

Key words: Malformations of cortical development, High-throughput nucleotide sequencing, Next generation sequencing, Animal disease models.

Introduction

Malformations of cortical development (MCD) include various structural abnormalities of the cerebral cortex that arise during formation of the cortical plate [1]. Development of the cerebral cortex is complex and involves proliferation, migration, and organization of neural progenitor cells [2,3]. Abnormalities at any stage during cerebral cortex development could result in a wide range of phenotypes, ranging from focal cortical dysplasia to hemimegalencephaly (HME) which show the enlargement of one hemisphere of brain (Fig. 1). Traditional classification of MCD has mainly been based on which developmental process

is likely to be affected. As novel genes and variants associated with MCD are discovered using next-generation sequencing (NGS) technology, an updated classification system for MCD has become more reliant on NGS findings [1,4]. Identification of disease-causing mutations, including inherited or *de novo* mutation, in MCD has revealed the molecular genetic basis of clinical manifestations [5-9]. As an example, recently identified genes responsible for microcephaly were found to encode DNA repair systems and centrosomal proteins that are crucial for regulating cell division [5,10,11].

The current challenge is to validate the increasing number of disease-causing mutations that are being identified by NGS.

Received: 16 December 2014, Revised: 18 March 2015, Accepted: 13 April 2015, Published: 30 June 2015

*Corresponding author: Jeong Ho Lee, M.D., Ph.D.

Graduate School of Medical Science and Engineering, KAIST, 291 Daehak-ro, Yuseong-gu, Daejeon 305-338, Korea.

Tel: +82-42-350-4246, Fax: +82-42-350-4240, E-mail: jhlee4246@kaist.ac.kr

Conflict of interest: We declare that we do not have any conflicts of interests.

© This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

© Copyright 2015 by the Korean Society of Medical Genetics

www.e-kjgm.org

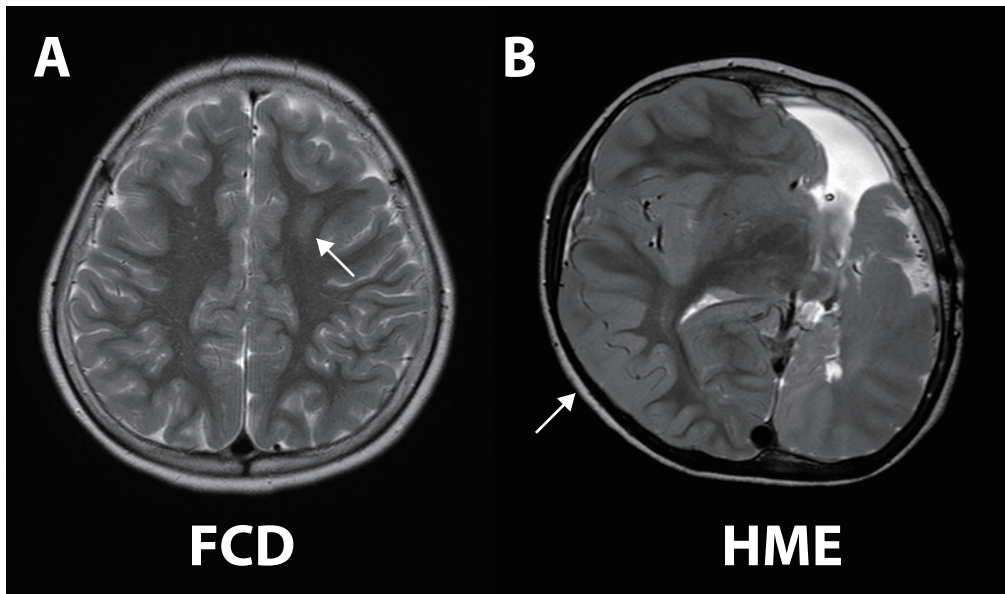


Fig. 1. Axial T2-weighted brain magnetic resonance imaging (MRI) of malformations of cortical development (MCD). (A) Brain MRI (axial view) of a focal cortical dysplasia (FCD) patient showing changes in white matter signals. Arrow highlight the focally affected cortical regions. (B) Brain MRI of a hemimegalencephaly (HME) patient showing enlargement of one cerebral hemisphere. Arrow indicate the affected hemispheres.

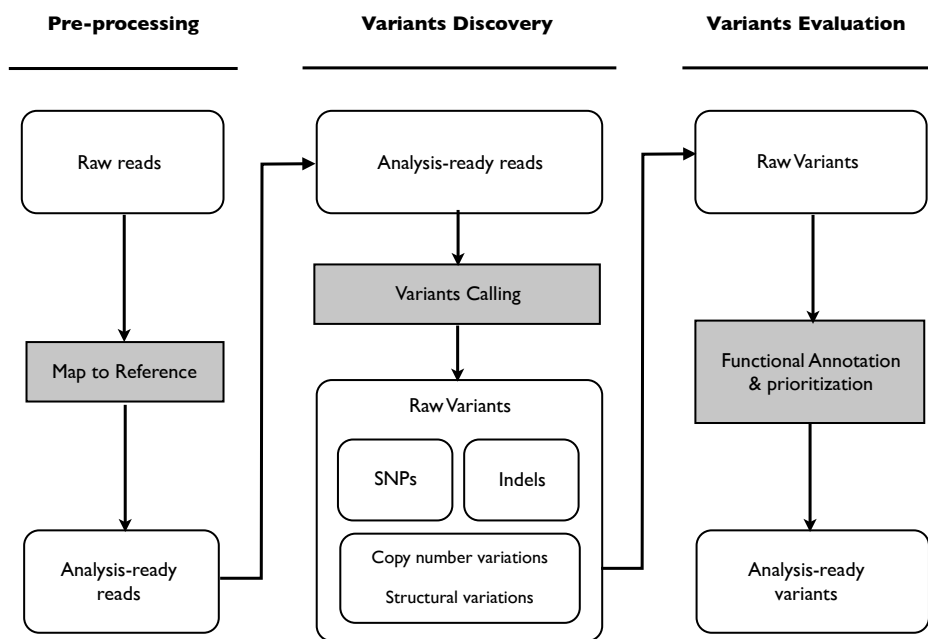


Fig. 2. Best practices, as stipulated by the Broad Institute for next-generation sequencing. Raw read data mapped to the correct genomic sequence and analysis-ready reads were analyzed to determine alternative alleles, including single nucleotide polymorphisms (SNPs), insertions/deletions (indels), copy number variation, and structural variation. 'Raw Variants', called during 'Variants Discovery', were annotated and selected as 'Analysis-ready variants'.

Although the identification of disease-associated mutations allows us to elucidate the molecular mechanisms of MCD, our understanding of the biological and pathological functions of the identified mutations is unclear. This has resulted in difficulties directly linking these disease-associated mutations to MCD. Genetic studies without disease modeling are not sufficient for revealing the causal relationship between genes and disease phenotypes. Disease modeling in combination with genetic studies and functional analyses at the organism level are required for elucidating causal relationships [12]. In this review, we describe the identification of novel genes associated with

MCD using NGS, and discuss how these identified genes can be functionally validated by *in vivo* disease modeling.

NGS-based Identification of Novel Genes

The advent of NGS technologies has significantly altered and accelerated research into the human genome. Among the various types of NGS platforms, whole-genome sequencing (WGS) can detect most genetic variations, such as single nucleotide variants, insertions/deletions, and copy number variation and rearrangement [13]. While the entire genome is

sequenced in WGS, whole-exome sequencing (WES) enriches the ~1% of genomic DNA that encodes proteins. Because of issues associated with data volume and cost effectiveness, WES is the more popular sequencing platform for identifying disease-causing mutations [12]. In the WES platform, the coding regions of entire genome are captured from sheared genomic DNA fragments, using labeled hybridization probes that bind to the protein-coding sequences [14]. The raw sequencing data from WES can be processed using well-established pipelines, such as Best Practices stipulated by the Broad Institute, that encompass pre-processing, variant discovery, and evaluation of variants [15,16] (Fig. 2). A fundamental limitation of WES is that it is unable to assess non-coding regions; however, many disease-causing genes involved in cortical developmental disorders have been identified using this technology [12,17]. In the following paragraph, we discuss specific approaches for detecting disease-causing mutations, including inherited and *de novo* somatic mutations.

1. Detection of inherited mutations

Prior to the development of NGS technologies, linkage analysis was mainly used to identify inherited mutations. Linkage analysis requires a large number of affected and unaffected individuals, spanning multiple generations [17]. The recent advent of NGS has altered how we identify inherited mutations in Mendelian disorders. NGS-based procedures require a smaller number of families than for linkage analysis. In addition, the raw sequencing data generated during NGS can be easily and rapidly processed using established protocols. As these tools for identifying mutations have developed, experimental design has become more important than sequencing itself with respect to studying inherited mutations.

To identify the mode of inheritance for mutations (autosomal recessive, autosomal dominant, or sex chromosome linked mutation), describing the family structure represented by pedigrees is essential. Family pedigree influences the number and selection of individuals to be sequenced, along with the data analysis methods employed. NGS-based identification of heterozygous, and homozygous mutations in autosomal recessive disorders is relatively straightforward. Homozygous variants in affected siblings and heterozygous variants in unaffected parents can be predicted; therefore, searching for overlapping mutations between siblings and parents (carrier) could determine disease-associated mutations. Using this strategy, several disease-causing genes have been identified in MCD [5–7,11,18,19]. A recent study regarding occipital cortical

gyration abnormalities, which is characterized by the presence of pachygyria and polymicrogyria, identified autosomal recessive *LAMC3* mutations [18]. In consanguineous families, the same heterozygous mutation found in both parents is commonly inherited in affected individuals. However, in a family without consanguinity, compound heterozygous mutations should be considered [12]. A recent analysis of non-consanguineous families with polymicrogyria, exhibiting shallow sulci and abnormal lamination, revealed a compound heterozygous mutation in the *WDR62* gene. This gene is a well-known disease-causing gene for a wide spectrum of cerebral cortical malformations, including microcephaly, lissencephaly, and pachygyria [19].

NGS-based detection of disease-causing genes in autosomal dominant diseases has been more challenging. For these diseases, it is not possible to rely on the homozygosity of an allele to detect disease-causing genes. Thus, it should be more careful to select one or a few disease-associated variants from a large number of candidate heterozygote variants [17]. The analysis of this type of mutations may stay on the present state for a while.

2. Detection of *de novo* somatic mutations

There is an increasing amount of evidence to support the importance of *de novo* mutations in neurodevelopmental diseases other than cancer [8,9,20,21]. During mitosis, *de novo* somatic mutations can occur without affecting cellular proliferation. Somatic mutations can occur at any stage during development with the frequency of mutated alleles varying. This is dependent on the timing of the mutation, which is consistent with mosaicism [22]. Because somatic mutations only exist in affected tissues, a simple comparison of non-overlapping mutations between affected and unaffected tissues is able to detect disease-associated mutations. Using this strategy, researchers have elucidated basic pathways related to cortical malformation [8,9,20,21]. A recent study of HME led to the identification of brain-specific somatic activating mutations in the *AKT3* (V-akt murine thymoma viral oncogene homolog 3) gene, which encodes serine-threonine kinase. *AKT3* is located upstream from the gene *MTOR* (mechanistic target of rapamycin), which is highly expressed in the developing brain during corticogenesis [20]. In addition, HME can also be caused by somatic mutations in other genes of the same pathway, including *PIK3CA* (phosphatidylinositol-4,5-bisphosphate 3-kinase, catalytic subunit alpha) and *MTOR* [8]. These studies suggest that brain overgrowth can result from mutations in

specific pathways that are responsible for the regulation of growth.

NGS analysis for somatic mutations consists of three steps: pre-processing; variant discovery; and variant evaluation (Fig. 2). Although there exists a general consensus regarding pre-processing protocols for preparing analysis-ready files from raw data, the numerous variant discovery protocols in use remain controversial. Therefore, various types of mutation callers are used depending on the type of mutation [23]. Each algorithm is based on different probability models; the variants reproducibly called by a number of independent algorithms are likely to be true variants rather than those called using a single algorithm. Therefore, the use of multiple detectors is required for improving the accuracy of somatic mutation detection [24]. In a recent study, it was shown that a consensus approach using multiple detectors provided high-confidence putative variants, by removing false positive predictions in each algorithm [25]. However, the further research into the optimum combinations of algorithms is required.

Although a multiple detector strategy is able to call true variants, many of the called variants are non-pathogenic. To select variants for functional studies among those that are called, prioritization of variants is essential. Various scoring systems, such as polyphen, scale-invariant feature transform and mutation assessor can be used to prioritize causal variants to increase filtering efficiency [26–28]. However, each annotation method contains parameters that make it difficult to estimate the relative importance of an independent annotation. To resolve these limitations, a general framework for integrating these various scores into a single value was recently introduced [29]. Despite these efforts, the prioritization of variants cannot fully resolve their biological functions in the context of diseases.

Functional Validation of Identified Variants

The major goal of genetic studies is to reveal the underlying pathogenesis of a disease and gain insights into the biological relevance of the relationship between genes and disease phenotypes. Although *in vitro* studies can help elucidate the molecular functions of identified variants, the *in vitro* environment does not necessarily recapitulate the physiological environment of an organism. To understand the biological function of identified variants in an organism, animal models carrying the disease-causing mutations are required.

1. Knockout and knock-in mouse models

Historically, mice have been widely used in laboratories as animal models. The rapid development of genetic engineering techniques since the 1980s has facilitated greater use of knockout (KO) and knock-in (KI) mice [30]. In KO mice, one or more target genes have been inactivated, while in KI mice disease-associated variant genes have been introduced into the mouse genome. Variants that confer a gain of function, or a loss of function, often result in alterations of the mouse phenotype including the appearance, behavior and molecular characteristics. As an example, in the tuberous sclerosis (TSC) KO mouse, the *Tsc1* or *Tsc2* genes are inactivated and mice exhibit many of the pathological characteristics of TSC, such as cytomegalic neuron, heterotopic neuron and astrocytosis [31,32]. In this particular model, the key neurological symptoms of TSC, such as intellectual disability and epilepsy, can be easily observed.

Another important genetic engineering technique, which was established in 1994, is the site-specific Cre-LoxP recombination system [33]. In this system, a specific target gene flanked by

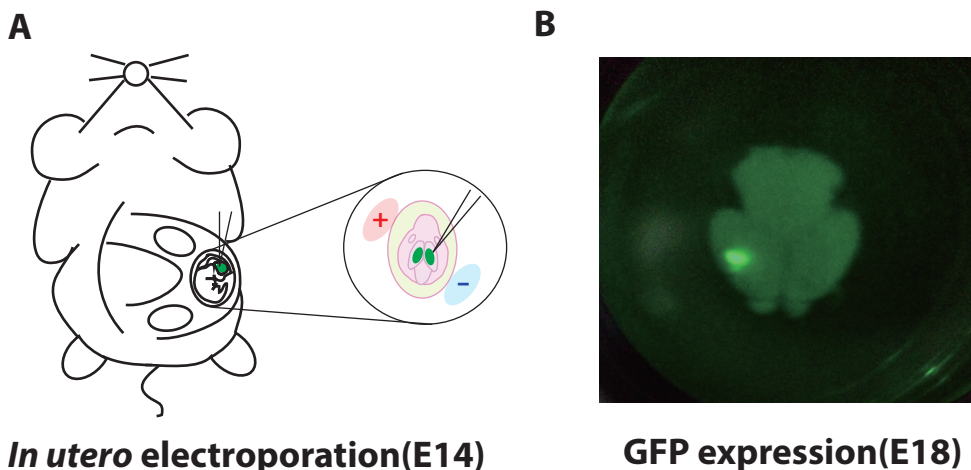


Fig. 3. *In utero* electroporation experiments. (A) Embryonic day 14 (E14) mouse embryos were subjected to *in utero* electroporation. Plasmid DNA with a green fluorescent protein (GFP) reporter was injected into one or both lateral ventricles and electrical pulses were applied to the brain from outside the uterine wall. (B) Focal expression of GFP after *in utero* electroporation. Brains were removed at embryonic day 18 (E18) and fixed; then, they were examined using a fluorescence stereomicroscope.

LoxP sequences is inactivated in a single organ or cell type, rather than the entire organism, through the region- or time-specific expression of Cre recombinase. This technique has helped elucidate cell-type or tissue-specific effects of disease-causing genes or mutations. Using Cre-LoxP recombination, a mouse overexpressing *LIS1* in the brain was generated; these mice had smaller brains, migration deficits, and exhibited loss of cellular polarity similar to that seen in lissencephaly patients [34]. However, there are some limitations to Cre-LoxP recombination, such as the maintenance of multiple lines, flanking gene effects, and embryonic lethality [35–37]. In addition, the lack of a specific promoter to control the expression of genes at a specific point in time, or in a particular region, limits the application of the mouse generated by Cre-loxP system [38]. Thus, the necessity of more efficient tools to control the gene expression arises in the research field of cortical development.

2. Brain somatic mutation mouse modeling

In recent years, *in utero* electroporation has emerged as an effective tool for manipulating gene expression in the brain at specific time points, and among particular cell populations [39] (Fig. 3). A target gene can be delivered to neural precursor cells, located in the ventricular zone, by injecting plasmid DNA into the mouse ventricle. Various sub-populations of neural progenitor cells can be transfected, depending on the stage of the embryo, or the location of where the electrical pulse is administered [40]. This technique is particularly useful for investigating basic mechanisms of cortical development, including migration and regional patterning [41,42]. Combining this technique with either RNA interference or Cre-LoxP recombination methods allows for both gain of function, and loss of function to be induced [43]. One of the most useful features of *in utero* electroporation is that it allows for focal cortical malformations to be recapitulated. Although *TSC1* conditional KO mice can be generated by crossing with Cre lines, they only mimic some of the features of TSC seen in patients such as seizure and dysmorphic neuron and do not show the heterotopic lesions containing the abnormal cells. In a recent study, discrete TSC-like mice were generated by *in utero* electroporation and were found to have heterotopic nodules characteristic of TSC [44]. Imprecise spatial control of gene expression, due to manually directing the electrode, is the most challenging problem of *in utero* electroporation. More accurate spatial control could be possible if plasmids carrying a specific promoter sequence were selectively expressed in target brain areas [45].

Conclusion

In the era of NGS, molecular genetic studies are rapidly shifting toward a more translational focus, with an attempt to bridge the gap between basic science and understanding the pathology of human diseases. Elucidating the genetic basis of MCD and generating appropriate mouse models have improved our understanding of the underlying mechanisms of these disorders and provided the opportunity to develop effective therapies [46]. The next subject in cortical development research will be to translate the large volume of genetic data for clinical applications.

Acknowledgements

This work was supported by a grant of the Korean Health Technology Research and Development (R&D) Project, Ministry of Health & Welfare, Republic of Korea (A121070 to J.H. Lee; H113C0208 to J.H. Lee), the Brain Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, Information and Communication Technology (ICT) & Future Planning (2013M3C7A1056564 to J.H. Lee).

References

1. Barkovich AJ, Guerrini R, Kuzniecky RI, Jackson GD, Dobyns WB. A developmental and genetic classification for malformations of cortical development: update 2012. *Brain* 2012;135:1348-69.
2. Pang T, Atefy R, Sheen V. Malformations of cortical development. *Neurologist* 2008;14:181-91.
3. Guerrini R. Genetic malformations of the cerebral cortex and epilepsy. *Epilepsia* 2005;46(Suppl)1:32-7.
4. Manzini MC, Walsh CA. What disorders of cortical development tell us about the cortex: one plus one does not always make two. *Curr Opin Genet Dev* 2011;21:333-9.
5. Bilgüvar K, Oztürk AK, Louvi A, Kwan KY, Choi M, Tatli B, et al. Whole-exome sequencing identifies recessive WDR62 mutations in severe brain malformations. *Nature* 2010;467:207-10.
6. Hussain MS, Baig SM, Neumann S, Nürnberg G, Farooq M, Ahmad I, et al. A truncating mutation of CEP135 causes primary microcephaly and disturbed centrosomal function. *Am J Hum Genet* 2012;90:871-8.
7. Poirier K, Lebrun N, Broix L, Tian G, Saillour Y, Boscheron C, et al. Mutations in TUBG1, DYNC1H1, KIF5C and KIF2A cause malformations of cortical development and microcephaly. *Nat Genet* 2013;45:639-

- 47.
8. Lee JH, Huynh M, Silhavy JL, Kim S, Dixon-Salazar T, Heiberg A, et al. De novo somatic mutations in components of the PI3K-AKT3-mTOR pathway cause hemimegalencephaly. *Nat Genet* 2012;44:941-5.
9. Rivière JB, Mirzaa GM, O'Roak BJ, Beddaoui M, Alcantara D, Conway RL, et al. De novo germline and postzygotic mutations in AKT3, PIK3R2 and PIK3CA cause a spectrum of related megalencephaly syndromes. *Nat Genet* 2012;44:934-40.
10. Yu TW, Mochida GH, Tischfield DJ, Sgaier SK, Flores-Sarnat L, Sergi CM, et al. Mutations in WDR62, encoding a centrosome-associated protein, cause microcephaly with simplified gyri and abnormal cortical architecture. *Nat Genet* 2010;42:1015-20.
11. Bakircioglu M, Carvalho OP, Khurshid M, Cox JJ, Tuysuz B, Barak T, et al. The essential role of centrosomal NDE1 in human cerebral cortex neurogenesis. *Am J Hum Genet* 2011;88:523-35.
12. Boycott KM, Vanstone MR, Bulman DE, MacKenzie AE. Rare-disease genetics in the era of next-generation sequencing: discovery to translation. *Nat Rev Genet* 2013;14:681-91.
13. Sims D, Sudbery I, Ilott NE, Heger A, Ponting CP. Sequencing depth and coverage: key considerations in genomic analyses. *Nat Rev Genet* 2014;15:121-32.
14. Bamshad MJ, Ng SB, Bigham AW, Tabor HK, Emond MJ, Nickerson DA, et al. Exome sequencing as a tool for Mendelian disease gene discovery. *Nat Rev Genet* 2011;12:745-55.
15. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytzky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010;20:1297-303.
16. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 2011;43:491-8.
17. Bras J, Guerreiro R, Hardy J. Use of next-generation sequencing and other whole-genome strategies to dissect neurological disease. *Nat Rev Neurosci* 2012;13:453-64.
18. Barak T, Kwan KY, Louvi A, Demirbilek V, Saygı S, Tüysüz B, et al. Recessive LAMC3 mutations cause malformations of occipital cortical development. *Nat Genet* 2011;43:590-4.
19. Murdock DR, Clark GD, Bainbridge MN, Newsham I, Wu YQ, Muzny DM, et al. Whole-exome sequencing identifies compound heterozygous mutations in WDR62 in siblings with recurrent polymicrogyria. *Am J Med Genet A* 2011;155A:2071-7.
20. Poduri A, Evrony GD, Cai X, Elhosary PC, Beroukhim R, Lehtinen MK, et al. Somatic activation of AKT3 causes hemispheric developmental brain malformations. *Neuron* 2012;74:41-8.
21. Jamuar SS, Lam AT, Kircher M, D'Gama AM, Wang J, Barry BJ, et al. Somatic mutations in cerebral cortical malformations. *N Engl J Med* 2014;371:733-43.
22. Poduri A, Evrony GD, Cai X, Walsh CA. Somatic mutation, genomic variation, and neurological disease. *Science* 2013;341:1237758.
23. Ding L, Wendl MC, McMichael JF, Raphael BJ. Expanding the computational toolbox for mining cancer genomes. *Nat Rev Genet* 2014;15:556-70.
24. Kim SY, Speed TP. Comparing somatic mutation-callers: beyond Venn diagrams. *BMC Bioinformatics* 2013;14:189.
25. Goode DL, Hunter SM, Doyle MA, Ma T, Rowley SM, Choong D, et al. A simple consensus approach improves somatic mutation prediction accuracy. *Genome Med* 2013;5:90.
26. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. *Nat Methods* 2010;7:248-9.
27. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res* 2003;31:3812-4.
28. Reva B, Antipin Y, Sander C. Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Res* 2011;39:e118.
29. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* 2014;46:310-5.
30. Thomas KR, Capecchi MR. Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. *Cell* 1987;51:503-12.
31. Uhlmann EJ, Wong M, Baldwin RL, Bajenaru ML, Onda H, Kwiatkowski DJ, et al. Astrocyte-specific TSC1 conditional knockout mice exhibit abnormal neuronal organization and seizures. *Ann Neurol* 2002;52:285-96.
32. Way SW, McKenna J 3rd, Mietzsch U, Reith RM, Wu HC, Gambello MJ. Loss of Tsc2 in radial glia models the brain pathology of tuberous sclerosis complex in the mouse. *Hum Mol Genet* 2009;18:1252-65.
33. Gu H, Marth JD, Orban PC, Mossmann H, Rajewsky K. Deletion of a DNA polymerase beta gene segment in T cells using cell type-specific gene targeting. *Science* 1994;265:103-6.
34. Bi W, Sapir T, Shchelochkov OA, Zhang F, Withers MA, Hunter JV, et al. Increased LIS1 expression affects human and mouse brain development. *Nat Genet* 2009;41:168-77.
35. Wolfer DP, Crusio WE, Lipp HP. Knockout mice: simple solutions to the problems of genetic background and flanking genes. *Trends Neurosci* 2002;25:336-40.
36. Crusio WE. Flanking gene and genetic background problems in genetically manipulated mice. *Biol Psychiatry* 2004;56:381-5.
37. Papaioannou VE, Behringer RR. Early embryonic lethality in genetically engineered mice: diagnosis and phenotypic analysis. *Vet Pathol* 2012;49:64-70.
38. Kratochwil CF, Rijli FM. The Cre/Lox system to assess the development

- of the mouse brain. *Methods Mol Biol* 2014;1082:295-313.
39. LoTurco J, Manent JB, Sidiqi F. New and improved tools for in utero electroporation studies of developing cerebral cortex. *Cereb Cortex* 2009;19(Suppl 1):i120-5.
 40. dal Maschio M, Ghezzi D, Bony G, Alabastri A, Deidda G, Brondi M, et al. High-performance and site-directed in utero electroporation by a triple-electrode probe. *Nat Commun* 2012;3:960.
 41. Takahashi M, Sato K, Nomura T, Osumi N. Manipulating gene expressions by electroporation in the developing brain of mammalian embryos. *Differentiation* 2002;70:155-62.
 42. Fukuchi-Shimogori T, Grove EA. Neocortex patterning by the secreted signaling molecule FGF8. *Science* 2001;294:1071-4.
 43. Bai J, Ramos RL, Ackman JB, Thomas AM, Lee RV, LoTurco JJ. RNAi reveals doublecortin is required for radial migration in rat neocortex. *Nat Neurosci* 2003;6:1277-83.
 44. Feliciano DM, Su T, Lopez J, Platel JC, Bordey A. Single-cell Tsc1 knockout during corticogenesis generates tuber-like lesions and reduces seizure threshold in mice. *J Clin Invest* 2011;121:1596-607.
 45. Shimogori T, Ogawa M. Gene application with in utero electroporation in mouse embryonic brain. *Dev Growth Differ* 2008;50:499-506.
 46. Beaulieu CL, Samuels ME, Ekins S, McMaster CR, Edwards AM, Krainer AR, et al. A generalizable pre-clinical research approach for orphan disease therapy. *Orphanet J Rare Dis* 2012;7:39.