



# Research Techniques Made Simple: Genome-Wide Homozygosity/Autozygosity Mapping Is a Powerful Tool for Identifying Candidate Genes in Autosomal Recessive Genetic Diseases

Hassan Vahidnezhad<sup>1,2,6</sup>, Leila Youssefian<sup>1,3,4,6</sup>, Ali Jazayeri<sup>5,6</sup> and Jouni Uitto<sup>1</sup>

Homozygosity mapping (HM), also known as autozygosity mapping, was originally used to map genes underlying homozygous autosomal recessive Mendelian diseases in patients from closely genetically related populations, followed by Sanger sequencing. With the increase in use of next-generation sequencing approaches, such as whole-exome sequencing and whole-genome sequencing, together with advanced bioinformatics filtering approaches, HM is again emerging as a powerful method for the identification of genes involved in disease etiology. In addition to its usefulness for research, HM is effective in clinical genetic services, increasing the efficiency of molecular diagnostics. For autosomal recessive Mendelian disorders with extensive genetic heterogeneity, HM can reduce both cost and turnaround time of mutation detection in the context of next-generation sequencing and can obviate expensive screening, such as biochemical testing in the setting of metabolic genodermatoses or antigen mapping for epidermolysis bullosa. It is therefore important for dermatology clinicians and researchers to understand the processes, principal uses, and advantages and limitations of HM when ordering or performing genetic tests for patients affected by heritable skin disorders.

*Journal of Investigative Dermatology* (2018) **138**, 1893–1900; doi:10.1016/j.jid.2018.06.170

**CME Activity Dates:** 21 August 2018

Expiration Date: 20 August 2019

Estimated Time to Complete: 1 hour

**Planning Committee/Speaker Disclosure:** All authors, planning committee members, CME committee members and staff involved with this activity as content validation reviewers have no financial relationships with commercial interests to disclose relative to the content of this CME activity.

**Commercial Support Acknowledgment:** This CME activity is supported by an educational grant from Lilly USA, LLC.

**Description:** This article, designed for dermatologists, residents, fellows, and related healthcare providers, seeks to reduce the growing divide between dermatology clinical practice and the basic science/current research methodologies on which many diagnostic and therapeutic advances are built.

**Objectives:** At the conclusion of this activity, learners should be better able to:

- Recognize the newest techniques in biomedical research.
- Describe how these techniques can be utilized and their limitations.
- Describe the potential impact of these techniques.

**CME Accreditation and Credit Designation:** This activity has been planned and implemented in accordance with the accreditation requirements and policies of the Accreditation Council for Continuing Medical Education through the joint providership of Beaumont Health and the Society for Investigative Dermatology. Beaumont Health is accredited by the ACCME to provide continuing medical education for physicians. Beaumont Health designates this enduring material for a maximum of 1.0 AMA PRA Category 1 Credit(s)<sup>™</sup>. Physicians should claim only the credit commensurate with the extent of their participation in the activity.

**Method of Physician Participation in Learning Process:** The content can be read from the Journal of Investigative Dermatology website: <http://www.jidonline.org/current>. Tests for CME credits may only be submitted online at <https://beaumont.cloud-cme.com/RTMS-Sept18> — click 'CME on Demand' and locate the article to complete the test. Fax or other copies will not be accepted. To receive credits, learners must review the CME accreditation information; view the entire article, complete the post-test with a minimum performance level of 60%; and complete the online evaluation form in order to claim CME credit. The CME credit code for this activity is: 21310. For questions about CME credit email [cme@beaumont.edu](mailto:cme@beaumont.edu).

<sup>1</sup>Department of Dermatology and Cutaneous Biology, Sidney Kimmel Medical College, Thomas Jefferson University, Philadelphia, Pennsylvania, USA;

<sup>2</sup>Biotechnology Research Center, Pasteur Institute of Iran, Tehran, Iran; <sup>3</sup>Department of Medical Genetics, Tehran University of Medical Sciences, Tehran, Iran;

<sup>4</sup>Genetics, Genomics and Cancer Biology PhD Program, Thomas Jefferson University, Philadelphia, PA, USA; and <sup>5</sup>Department of Information Science, College of Computing and Informatics, Drexel University Philadelphia, Pennsylvania, USA

<sup>6</sup>These authors contributed equally to this work.

Correspondence: Jouni Uitto, Department of Dermatology and Cutaneous Biology, Sidney Kimmel Medical College at Thomas Jefferson University, 233 South 10th Street, Suite 450 BLSB, Philadelphia, Pennsylvania 19107, USA. E-mail: [Jouni.Uitto@jefferson.edu](mailto:Jouni.Uitto@jefferson.edu)

Abbreviations: AR, autosomal recessive; EB, epidermolysis bullosa; GATK, Genome Analysis Toolkit; HM, homozygosity mapping; Mb, mega base pair; NGS, next-generation sequencing; ROH, region of homozygosity; SNP, single nucleotide polymorphism; WES, whole-exome sequencing; WGS, whole-genome sequencing

## BENEFITS

- HM reduces costs and turnaround time and increases the yield of mutation analysis by NGS.
- HM can obviate the need for specialized screening tests, such as biochemical testing in the setting of metabolic genodermatoses or antigen mapping for EB.
- HM can minimize the need for sequencing multiple genes in cases of genetically heterogeneous AR genodermatoses.
- HM can provide evidence for pathogenicity of previously unsuspected mutations, such as deep intronic or missense variants of uncertain significance.
- HM is a high-throughput, genome-wide method that can provide clues for discovering novel disease-related genes.

## LIMITATIONS

- HM is mostly applicable to patients born to consanguineous parents.
- HM is applicable only to Mendelian diseases with a homozygous AR pattern of inheritance.
- In rare cases, consanguineous families can carry compound heterozygous mutations, which will not be detectable in the region harboring the mutated gene.

## INTRODUCTION

Genetic skin diseases, or genodermatoses, are a large category of heritable single-gene (Mendelian) diseases, with over 1,000 genes currently being associated with cutaneous manifestations. About 50% of genetic skin diseases are inherited as autosomal recessive (AR) disorders. The routine diagnosis of genetic skin diseases is complicated by the fact that in this group of disorders, clinical manifestations may result from mutations in unrelated genes (genetic heterogeneity), and mutations in the same gene often lead to dissimilar clinical signs (phenotypic heterogeneity). Other compounding factors include the existence of new genes and/or presence of novel disease subtypes (Mizrachi-Koren et al., 2006).

For many years, homozygosity mapping (HM) was the primary tool used for genetic mapping of AR Mendelian disorders in patients from genetically closely related parents. However, with the recent increase in use of next-generation sequencing (NGS) approaches, such as whole-exome sequencing (WES) and whole-genome sequencing (WGS), together with advanced bioinformatics filtering approaches, HM is again emerging as a powerful method for identifying candidate genes involved in disease etiology (Ott et al., 2015). In addition to being a useful research tool, HM has been proven to be a useful adjunct in the practice of clinical

genetic services, which can dramatically reduce the cost and turnaround time for molecular diagnosis of homozygous AR genetic skin diseases (Alkuraya, 2010).

HM exploits the fact that patients born to consanguineous parents likely inherited two recessive copies of a mutant allele from a common ancestor. Because both alleles are the same and originated from a common ancestor, they are known as “identical-by-descent” alleles. Because the chromosomal regions tend to be transmitted intact, except for a few recombinational events in each generation, a patient born to consanguineous parents will have a chromosomal region flanking the disease locus in which the genetic markers are tandemly homozygous by descent (Figure 1). In addition, other identical-by-descent regions exist that are unrelated to the disease. The principle of the HM method is to search for genetic regions of homozygosity (ROHs), varying from a few to several megabase pairs (Mb), in patients’ DNA, followed by identification of the region that harbors a mutated gene involved in rare recessive traits. If more than one affected individual exists in the extended family, the strategy will be to look for ROHs that are exclusively shared by all affected individuals and not present in healthy close relatives, such as the parents. On rare occasions, the AR disorder even in a consanguineous family can be due to compound heterozygous mutations, and HM may not lead to detection of the region harboring the mutated gene (Alkuraya, 2012).

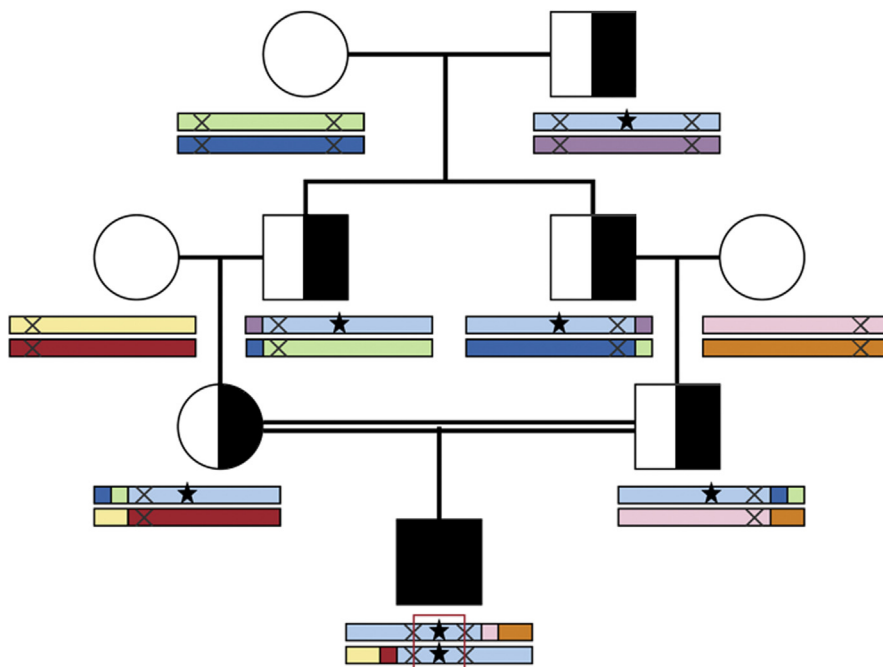
A number of screening methods, taking advantage of the extensive polymorphism of the human genome, have been shown to facilitate the identification of candidate genes. The advent of high-density single nucleotide polymorphism (SNP) arrays has allowed genome-wide mapping of ROHs at high resolution for a relatively low cost in consanguineous and outbred (e.g., in the case of a founder effect) families (Alkuraya, 2010; Ott et al., 2015; Schuurs-Hoeijmakers et al., 2011). HM can be used alone for mapping of causative genes before mutation analysis by Sanger sequencing or in combination with NGS to improve the mutation detection rate.

This synopsis will review the methodology and bioinformatics involved in HM and will provide a few examples of HM application. Although most dermatologists or laboratory researchers may not have the necessary computing skills or bioinformatics expertise to perform the technical aspects of WES and WGS, this review should provide guidelines for them to apply these techniques, which are commercially available.

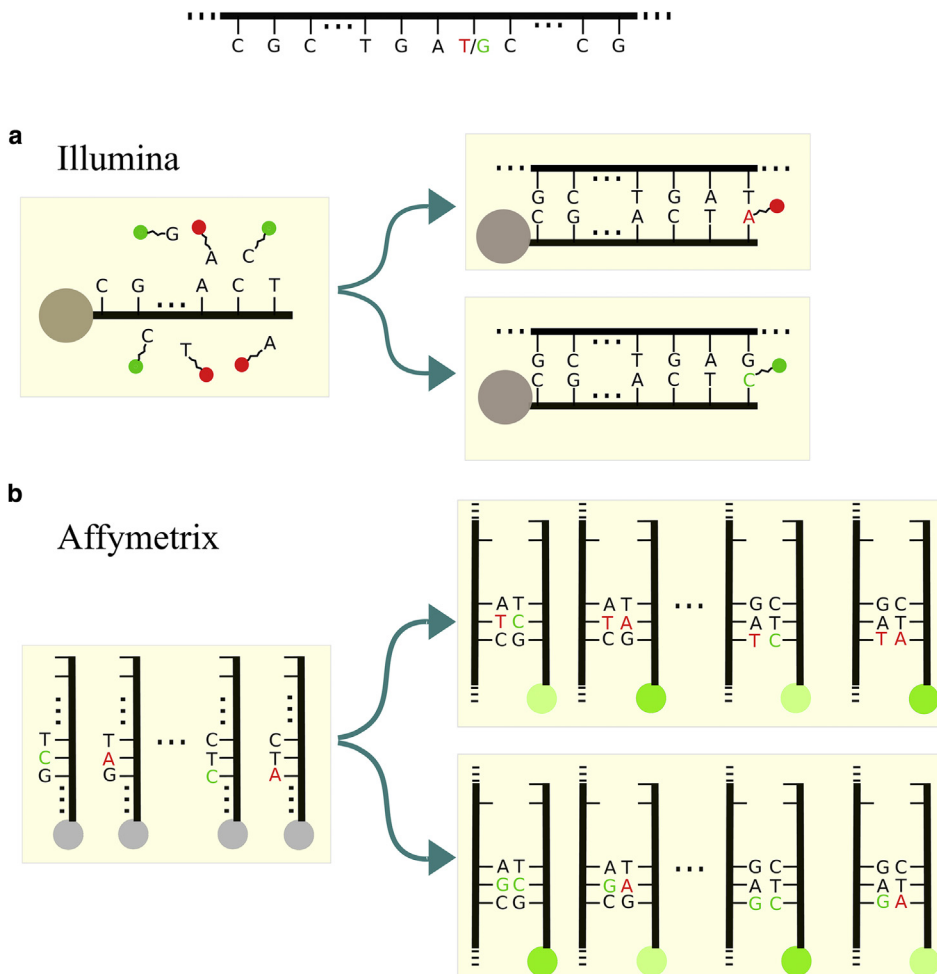
## METHODOLOGIES AND BIOINFORMATICS OF HM

### Chemistry of different platforms of SNP-based array

Genome-wide HM can be performed with widely used SNP-based array platforms developed by two competing companies, Affymetrix (Santa Clara, CA) and Illumina (San Diego, CA) (Figure 2). With the Affymetrix system, amplified and labeled DNA is hybridized to an array containing 25-mer oligonucleotides, each of which determines a specific SNP. Illumina platforms for SNP genotyping use oligonucleotides attached to silica beads as hybridization probes. The probe undergoes single-base extension with tagged terminating nucleotides. The extended fluorescent probes are then scanned on the BeadArray Reader (LaFramboise, 2009).



**Figure 1. Schematic representation of a pedigree with first cousin consanguineous parents who have two affected and one unaffected offspring.** An ancestral haplotype harboring a mutation (star) was tracked as it transferred to descendants. In each generation, different haplotypes were entered to the pedigree depicted with a specific color. Recombinational events (shown as crosses) in each generation shorten the size of the haplotype harboring the mutation. The recombinational events shown in the parents of patients occur only in the older patient.



**Figure 2. Overview of chemistry of different platforms of SNP array technology.** A fragment of DNA harboring a T/G SNP is shown at the top. (a) A 50-nt probe complementary to the sequence adjacent to the SNP site is attached to each Illumina (San Diego, CA) bead. After hybridization, a single-base extension (A or C) that is complementary to the allele carried by the DNA (T or G, respectively) yields an appropriately colored signal (red or green). (b) In the Affymetrix (Santa Clara, CA) system, there are several different probes with 25-nt length for each SNP. Each probe has a different nucleotide at a central position. Not only the central nucleotide but other nucleotides at positions +1, 0, and -1 relative to the central nucleotide can be variable in each probe. The DNA binds to probes regardless of the allele it carries, but it does so more efficiently when it is complementary to all 25 bases (lighter green) rather than mismatching the SNP site (stronger green). Thus, with increasing degree of complementarity, there is an increase in the brightness of the signal. A, adenine; C, cytosine; G, guanine; nt, nucleotide; SNP, single nucleotide polymorphism; T, thymine.

### Bioinformatics workflow

ROHs can be quickly detected from either SNP genotype array or WGS/WES data, and each step of data processing for HM is briefly described below. SNP array and WGS/WES output data are in the format of .ped or .map and fastq or uBAM, respectively (Figures 3 and 4). Depending on the type of data, the workflow of HM can be initiated from specific starting points. Figure 3 shows a panoramic view of the entire process, and a more finely detailed version of the process is shown in Figure 4.

### HM using WES/WGS data

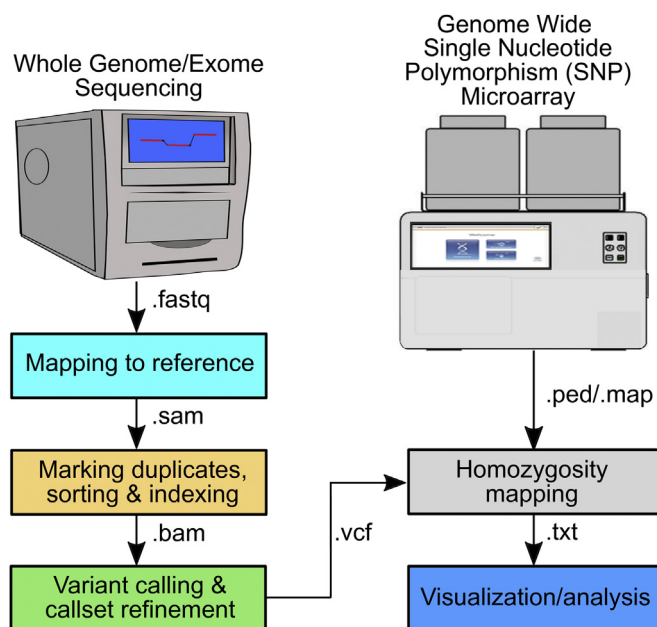
When workflow begins from fastq or uBAM files produced by WES/WGS, the first step of HM is mapping sequences to the reference genome (Li and Durbin, 2010). Mapped sequences are then sorted and converted to BAM files, and duplicates are marked. These steps can be performed using Picard (For details of software programs and on-line tools, see Table 1).

To improve the accuracy and quality of scores and to remove biases, the generated BAM files go through InDels (i.e., insertion/deletion) realignment and base quality score recalibration steps using Genome Analysis Toolkit (GATK), which is a software developed by the Broad Institute for variant discovery in high-throughput sequencing data. In these two steps, GATK first improves the original alignment of reads, and in the base quality score recalibration step, using machine learning approaches, GATK eliminates systematic errors generated by sequencing machines and adjusts base reads quality scores. The InDel realignment step is not required in cases in which variant callers with a reassembly step such as HaplotypeCaller are used. The output will be BAM files ready for variant calling or direct exploration and visualization using Integrative Genomics Viewer (IGV) tools (Robinson et al., 2011). In the latter case, the BAM files should be indexed; Picard and IGVtool modules can be used for this purpose.

The next step in the workflow of bioinformatics analysis is variant discovery, which is recommended by GATK to be performed jointly on members of the family, including patients and healthy relatives if data are available (McKenna et al., 2010). For this purpose, variants are called, and a Genomic Variant Call Format (GVCF) file is generated for each BAM file separately. All GVCF files are then genotyped jointly to create one VCF file. For the sake of accuracy, variants are recalibrated, assigning a new quality score to each variant call. The generated VCF file can be used directly for variant analysis purposes. In this case, GATK recommends additional steps, such as genotype refinement, variant annotation, and variant evaluation. The generated VCF files are then converted to binary/standard formats. In this case, PLINK (Purcell et al., 2007) provides commands to set the individual and family identifications and generate binary/standard formats for HM.

### HM using whole-genome SNP array

ROHs can also be quickly detected by SNP genotype array data with the format of .ped/.map. The SNP genotyping data can be analyzed using PLINK. The PLINK default values are appropriate for finding large segments of ROHs present on dense genotyping platforms and can be left unchanged during



**Figure 3. A panoramic view of the process of bioinformatics analysis for detection of ROHs from either SNP genotype array or WGS/WES data.**

SNP array and WGS/WES output data are in the format of .ped or .map and fastq or uBAM, respectively. Depending on the type of data, the workflow of HM can be initiated from specific starting points. Color coding is identical in Figures 3 and 4. HM, homozygous mapping; SNP, single nucleotide polymorphism.

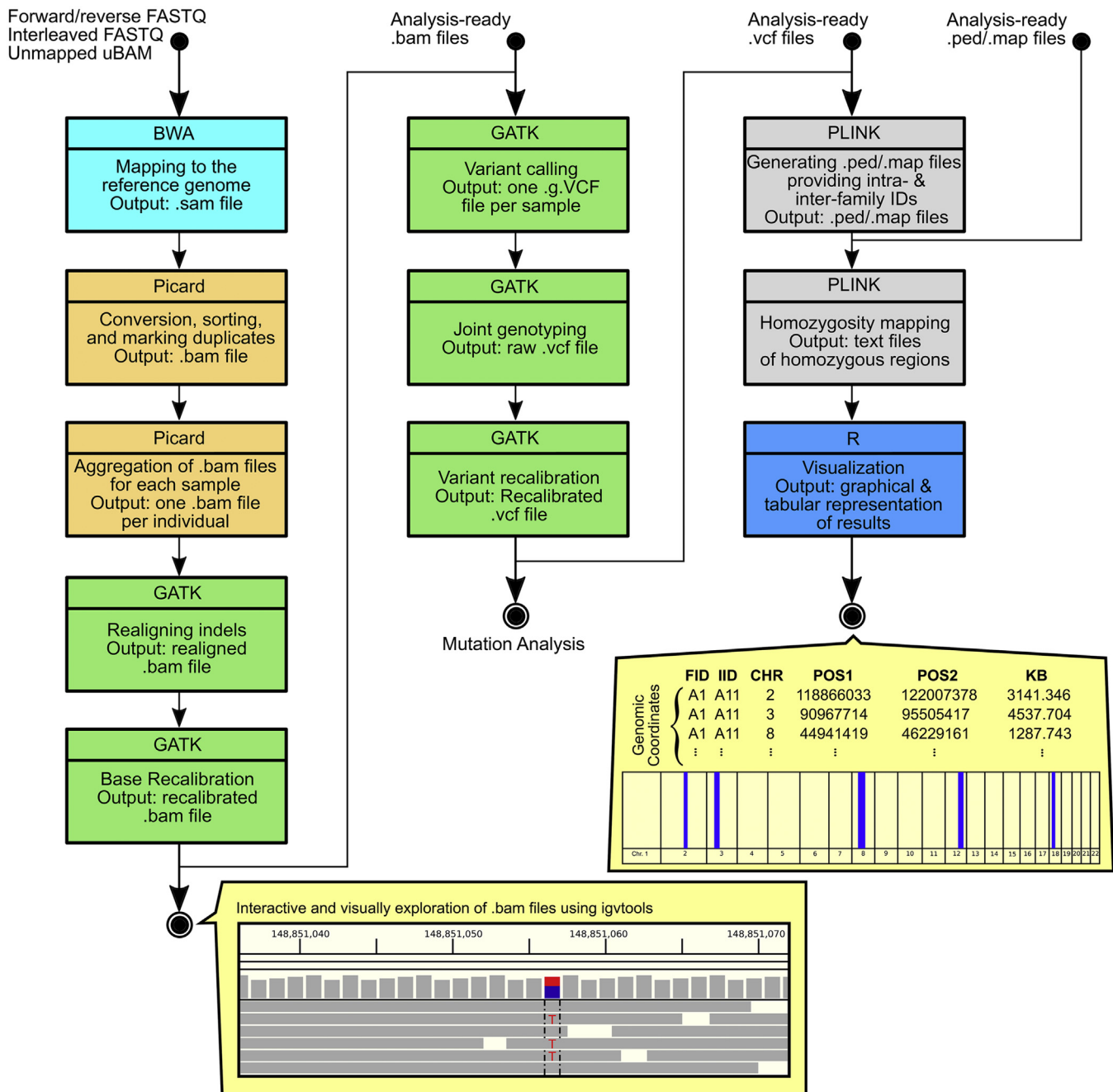
the analysis. PLINK also provides filtration commands to eliminate ROHs with predefined thresholds (e.g., <2 Mb). For identification of ROHs, both the graphical user interface provided by PLINK (gPLINK) and the command-line interface can be used. The results can be in the format of start and end positions of shared ROHs among all patients within a family, known as genomic coordinates (Figure 4). For each person, we expect around 10 genomic coordinates belonging to different chromosomal regions. PLINK is not the only available tool for HM. Other tools, such as GERMLINE (Gusev et al., 2009) and HomSI (Gomez et al., 2014), can also be used for this purpose. However, PLINK is recommended because of the graphical user interface (gPLINK) and extensive online documentation provided and because it is free, open source, and easy to download and run. PLINK is compatible with the data generated by the other tools in previous steps of workflow of HM discussed in this review, or it provides the conversion commands to produce the eligible formats. The output can be tabulated and visualized using different tools, such as R (<https://www.r-project.org/>).

### The Genomic Oligoarray and SNP Array Evaluation Tool

The sequence variants identified by sequencing with the assistance of HM can be classified as nonsense, missense, splice site, or indel, and can be identified as deleterious to the gene/protein function by Annotate Variation (ANNOVAR) scores. In support of pathogenicity, their minor allelic frequencies can be determined in ExAC and 1000 Genomes databases, coupled with segregation analysis in the families.

The Genomic Oligoarray and SNP Array Evaluation Tool (version 3.0) is an online tool to accelerate and improve





**Figure 4. Detailed version of the process of bioinformatics analysis for detection of ROHs from either SNP genotype array or WGS/WES data.** Each box in the figure shows the tool used for the corresponding step, the process description, and the output files. These steps are excerpted from literature and best practices recommended by Burrows-Wheeler Aligner (BWA) (Table 1; Li and Durbin, 2010; McKenna et al., 2010; Purcell et al., 2007). Color coding is identical in Figures 3 and 4. The usual outputs of each method are depicted. GATK, Genome Analysis Toolkit; ROH, region of homozygosity; SNP, single nucleotide polymorphism; WES, whole-exome sequencing; WGS, whole-genome sequencing.

clinical interpretation of SNP array results for diagnostic purposes in cases of close familial genetic relationships. This Web-based program permits submission of ROHs as genomic coordinates and retrieves genes within these regions and their associated AR disorders using built-in Online Mendelian Inheritance in Man (OMIM), University of California–Santa Cruz (UCSC), and National Center for Biotechnology Information (NCBI) databases. It allows the user to further filter to generate a short list of candidate conditions relevant for the diagnosis,

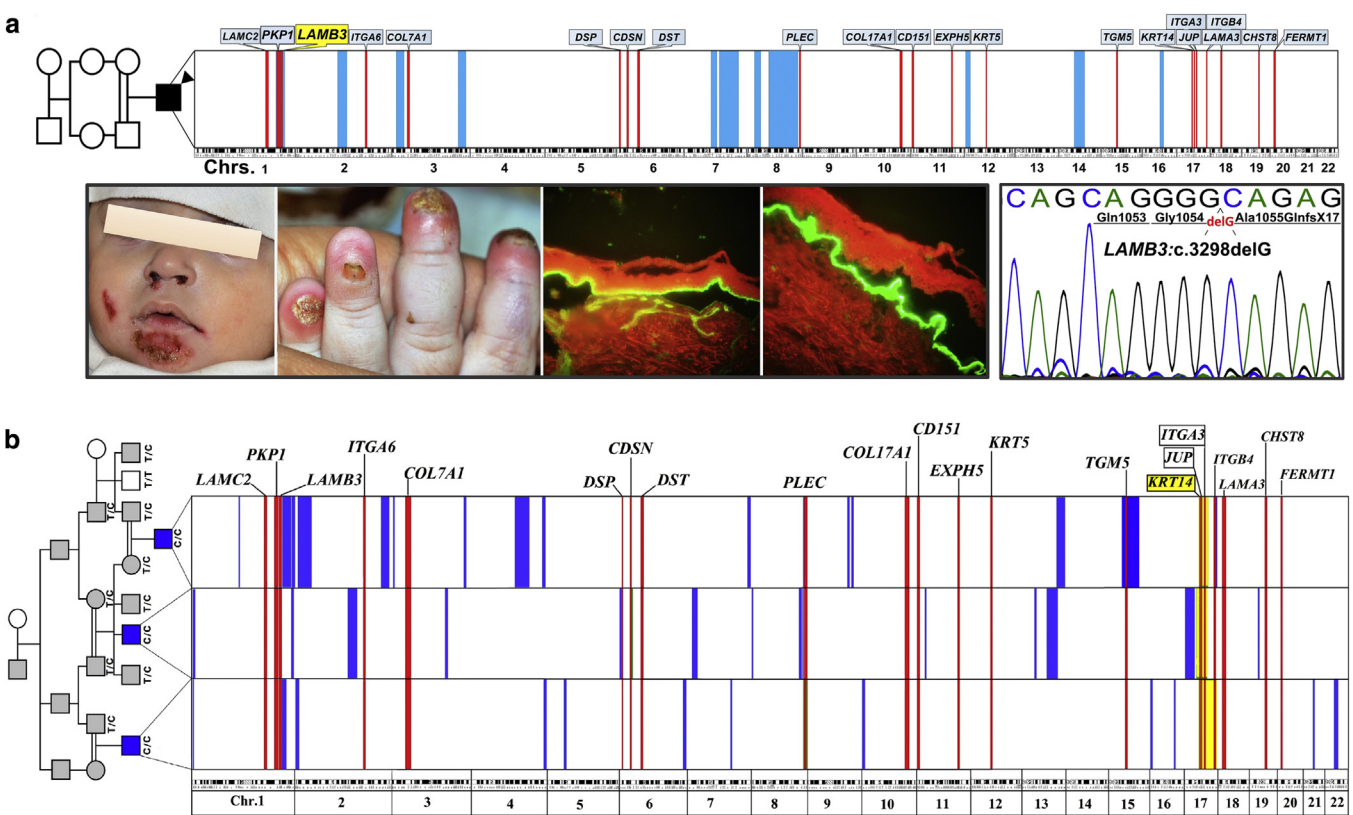
making it possible to strategize a focused diagnostic testing approach. Relevant OMIM clinical synopses can be submitted with key clinical terms, permitting further filtering for candidate genes and disorders (Wierenga et al., 2013).

#### EXAMPLES OF UTILITY OF HM IN EPIDERMOLYSIS BULLOSA PATIENTS

We have investigated the utility of HM for the molecular diagnosis of heterogeneous AR disorders, using unknown

Software	Description and Purpose	URL	References
BWA (Burrows-Wheeler Aligner)	Package used to map read sequences to a reference genome	<a href="http://bio-bwa.sourceforge.net/">http://bio-bwa.sourceforge.net/</a>	Li and Durbin (2009, 2010)
GATK (Genome Analysis Toolkit)	A multipurpose variant discovery and genotyping tool	<a href="https://software.broadinstitute.org/gatk/">https://software.broadinstitute.org/gatk/</a>	McKenna et al. (2010)
Picard	Tools used for manipulating sequencing data in different formats such as BAM and VCF files	<a href="http://broadinstitute.github.io/picard">http://broadinstitute.github.io/picard</a>	Refer to corresponding URL
PLINK	Tools used for whole genome association studies. It can be used for homozygosity mapping and identical-by-descent estimation	<a href="http://zzz.bwh.harvard.edu/plink/">http://zzz.bwh.harvard.edu/plink/</a>	Purcell et al. (2007)
R	A Software environment for statistical data analysis and visualization	<a href="https://www.r-project.org/">https://www.r-project.org/</a>	Refer to corresponding URL
SAMtools	Tools can be used for manipulating SAM/BAM formats	<a href="http://samtools.sourceforge.net/">http://samtools.sourceforge.net/</a>	Li et al. (2009)
Genomic Oligoarray and SNP Array Evaluation Tool version 3.0	Web-based tool developed for retrieving genes and their associated autosomal recessive disorders within regions of homozygosity	<a href="http://firefly.ccs.miami.edu/cgi-bin/ROH/ROH_analysis_tool.cgi">http://firefly.ccs.miami.edu/cgi-bin/ROH/ROH_analysis_tool.cgi</a>	Wierenga et al. (2013)

Abbreviation: SNP, single nucleotide polymorphism.



**Figure 5. Utility of HM for the molecular diagnosis of heterogeneous AR disorders using unknown types of EB as a paradigm.** (a) Homozygosity mapping, representative clinical features, immune-epitope mapping, and mutation analysis of the candidate gene in a case of EB, originally diagnosed as laryngo-onycho-cutaneous (Shabbir) syndrome. Note the cutaneous erosions and excess granulation tissue on the chin and nasal cavity and dystrophic changes in the nails consistent with diagnosis of laryngo-onycho-cutaneous syndrome. An Illumina (San Diego, CA) SNP panel of 240,000 markers was used to identify homozygosity blocks of 2 Mb or greater (vertical blue lines) along the entire autosome; chromosomes 1–22 are listed at the bottom. The genomic loci of candidate genes known to be associated with EB are indicated by vertical red lines. Only one EB-related gene co-aligned with a homozygosity block, implicating *LAMB3* on chromosome region 1q32.2 (yellow box). Staining with a monoclonal antibody against integrin  $\beta 4$  marked the blister roof (left) and with a monoclonal antibody against collagen VII marked the blister floor (right); localization of these two proteins in immunofluorescence analysis indicated that the level of cleavage is within the lamina lucida, suggesting the diagnosis of junctional EB. Sanger sequencing showed the mutation c.3163delG (p.Ala1055Glnfs\*17) in the *LAMB3* gene (for details, see Vahidnezhad et al., 2018b). (b) HM in three affected members of a family with severe generalized blistering in EB simplex. In this family, only one homozygosity block (blue lines) shared by all three affected individuals was noted on chromosome region 17q21 (yellow lines). This interval harbors three EB-associated genes, *KRT14*, *JUP*, and *ITGA3*. Sequence analysis showed a p.Ile377Thr mutation in *KRT14* co-segregating with the phenotype in a semidominant pattern (for details, see Vahidnezhad et al., 2016). AR, autosomal recessive; EB, epidermolysis bullosa; HM, homozygosity mapping; Mb, megabase.

types of epidermolysis bullosa (EB) as a paradigm (Vahidnezhad et al., 2016, 2017b, 2018b). EB is caused by mutations in as many as 20 genes, and identification of specific mutations is critical for molecular confirmation of the diagnosis and precise subclassification with prognostic implications (Has et al., 2018; Uitto et al., 2016; Vahidnezhad et al., 2018a).

HM significantly facilitates the identification of the candidate genes in several different ways. For example, routine diagnosis of EB entails invasive antigen mapping of skin biopsy to show the level of cleavage and lack of mutated protein, which is being expertly performed in a few laboratories around the world. In addition, antigen mapping costs several hundred dollars. The results of antigen mapping provide clues for candidate genes for Sanger sequencing. However, if only one gene was identified by co-alignment of the putative candidate gene loci and the homozygosity blocks in HM, this would allow us to focus on characterization of a single gene, with considerable savings of cost and effort (the current in-house price for HM is \$50 for 650,000 SNP markers). Furthermore, HM of additional affected members of the family, besides the proband, facilitates identification of candidate genes by NGS, allowing prioritization of the analysis by bioinformatics. For example, for each patient we typically have approximately 100,000 variants using WES data. The number of variants that should be examined for detection of disease-causing genes is reduced to approximately 10,000 by using HM, which shows a significant decrease in required resources. For example, in case 1 (Figure 5a), the patient was born to consanguineous first cousin parents and was the only EB patient in the family; the initial diagnosis was laryngo-onycho-cutaneous (Shabbir) syndrome, a type of EB usually caused by mutations in *LAMA3*. However, gene-targeted sequencing of *LAMA3* in this patient did not identify a mutation in this gene. Genome-wide HM identified 11 ROHs of 2 Mb or greater, and alignment of the positions of the 20 EB-associated genes suggested *LAMB3* as the only candidate gene in the proband. Subsequent Sanger sequencing showed a homozygous mutation in *LAMB3*:c.3298delG (Figure 5a). This case illustrates the success of our current approaches to reach the correct diagnosis of generalized junctional EB (Vahidnezhad et al., 2017a, 2018b).

Case 2 is an example of HM for an extended family with more than one patient. HM in three patients of this pedigree showed a number of ROHs in each patient and a homozygosity block shared by all three affected individuals on chromosomal region 17q21 corresponding to *KRT14*. An EB candidate gene was therefore identified (Figure 5b). Sequencing of *KRT14* showed a homozygous c.1130T>C, p.Ile377Thr mutation in the three individuals (Vahidnezhad et al., 2016).

### CONCLUSIONS AND FUTURE DIRECTIONS

HM is a powerful method for gene mapping, particularly of homozygous AR Mendelian diseases, in research and clinical settings. With the recent increase in use of NGS approaches, HM has been a powerful method for streamlining the identification of mutated genes in conjunction with NGS bioinformatics filtering approaches. Thus, HM can significantly

### MULTIPLE CHOICE QUESTIONS

- Whole genome homozygosity mapping is applicable for which of the following?
  - Autosomal dominant Mendelian disorders
  - Autosomal recessive Mendelian disorders
  - X-linked Mendelian disorders
  - Genetic disorders with mitochondrial inheritance
- Which of the following is NOT a high-throughput (genome-wide) method for homozygosity mapping?
  - Whole-genome sequencing
  - Whole-exome sequencing
  - Short tandem repeat (STR) genotyping
  - Single nucleotide polymorphism (SNP) array with 240,000 markers
- Homozygosity mapping can map genes for which type of mutations?
  - Autosomal recessive compound heterozygous mutations
  - Autosomal recessive homozygous mutations
  - X-linked recessive mutations
  - Dominant negative mutations
- Which of the following methods can map a causal gene and detect its underlying mutation?
  - Genome-wide single nucleotide polymorphism (SNP) array with Illumina platform
  - Whole-exome sequencing (WES)
  - Short tandem repeat (STR) genotyping
  - Genome-wide single nucleotide polymorphism (SNP) array with Affymetrix platform
- Which of the software packages is able to find regions of homozygosity (ROHs) from .vcf and .ped/.map file formats?
  - PLINK
  - BWA
  - GATK
  - Picard

reduce the cost and turnaround time of mutation detection and obviate the need for extensive screening tests, thus increasing the efficiency of molecular diagnostics. In addition, HM can provide evidence that previously unsuspected mutations are pathogenic, such as deep intronic or missense variants of uncertain significance. An improvement in HM could consist of WGS that would allow fine-mapping and sequencing of the mutant genes in rare genetic diseases. Current limitations of this approach include the relatively high

cost of WGS and the requirement for high-capacity computer systems for data analysis. However, these limitations are expected to be overcome with advanced technologies in the near future.

## ORCID

Hassan Vahidnezhad: <http://orcid.org/0000-0003-4298-9147>

Leila Youssefian: <http://orcid.org/0000-0002-4253-6503>

Ali Jazayeri: <http://orcid.org/0000-0002-6468-1971>

Jouni Uitto: <http://orcid.org/0000-0003-4639-807X>

## CONFLICT OF INTEREST

The authors state no conflict of interest.

## ACKNOWLEDGMENTS

Carol Kelly assisted in manuscript preparation. The original research by the authors was supported in part by DEBRA International. This study is in partial fulfillment of the PhD thesis of HV.

## SUPPLEMENTARY MATERIAL

Supplementary material is linked to this paper. Teaching slides are available as supplementary material.

## REFERENCES

- Alkuraya FS. Homozygosity mapping: one more tool in the clinical geneticist's toolbox. *Genet Med* 2010;12:236–9.
- Alkuraya FS. Discovery of rare homozygous mutations from studies of consanguineous pedigrees. *Curr Protoc Hum Genet* 2012;75:6.12.1–13.
- Gormez Z, Bakir-Gungor B, Sagioglu MS. HomSI: a homozygous stretch identifier from next-generation sequencing data. *Bioinformatics* 2014;30:445–7.
- Gusev A, Lowe JK, Stoffel M, Daly MJ, Altshuler D, Breslow JL, et al. Whole population, genome-wide mapping of hidden relatedness. *Genome Res* 2009;19:318–26.
- Has C, Nystrom A, Saeidian AH, Bruckner-Tuderman L, Uitto J. Epidermolysis bullosa: molecular pathology of connective tissue components in the cutaneous basement membrane zone [e-pub ahead of print]. *Matrix Biol* 2018; <https://doi.org/10.1016/j.matbio.2018.04.001> (accessed 25 July 2018).
- LaFramboise T. Single nucleotide polymorphism arrays: a decade of biological, computational and technological advances. *Nucleic Acids Res* 2009;37:4181–93.
- Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009;25:1754–60.
- Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 2010;26:589–95.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009;25:2078–9.
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernysky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010;20:1297–303.
- Mizrachi-Koren M, Shemer S, Morgan M, Indelman M, Khamaysi Z, Petronius D, et al. Homozygosity mapping as a screening tool for the molecular diagnosis of hereditary skin diseases in consanguineous populations. *J Am Acad Dermatol* 2006;55:393–401.
- Ott J, Wang J, Leal SM. Genetic linkage analysis in the age of whole-genome sequencing. *Nat Rev Genet* 2015;16:275–84.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007;81:559–75.
- Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. *Nat Biotechnol* 2011;29:24–6.
- Schuurs-Hoeijmakers JH, Hehir-Kwa JY, Pfundt R, van Bon BW, de Leeuw N, Kleefstra T, et al. Homozygosity mapping in outbred families with mental retardation. *Eur J Hum Genet* 2011;19:597–601.
- Uitto J, Vahidnezhad H, Youssefian L. Genotypic heterogeneity and the mode of inheritance in epidermolysis bullosa. *JAMA Dermatol* 2016;152:517–20.
- Vahidnezhad H, Youssefian L, Saeidian AH, Mozafari N, Barzegar M, Sotoudeh S, et al. *KRT5* and *KRT14* mutations in epidermolysis bullosa simplex with phenotypic heterogeneity, and evidence of semidominant inheritance in a multiplex family. *J Invest Dermatol* 2016;136:1897–901.
- Vahidnezhad H, Youssefian L, Saeidian AH, Touati A, Sotoudeh S, Abiri M, et al. Multigene next generation sequencing panel identifies pathogenic variants in patients with unknown subtype of epidermolysis bullosa: subclassification with prognostic implications. *J Invest Dermatol* 2017a;137:2649–52.
- Vahidnezhad H, Youssefian L, Zeinali S, Saeidian AH, Sotoudeh S, Mozafari N, et al. Dystrophic epidermolysis bullosa: *COL7A1* mutation landscape in a multi-ethnic cohort of 152 extended families with high degree of customary consanguineous marriages. *J Invest Dermatol* 2017b;137:660–9.
- Vahidnezhad H, Youssefian L, Saeidian AH, Mahmoudi HR, Touati A, Abiri M, et al. Recessive mutation in tetraspanin *CD151* causes Kindler syndrome-like epidermolysis bullosa with multi-systemic manifestations including nephropathy. *Matrix Biol* 2018a;66:22–33.
- Vahidnezhad H, Youssefian L, Saeidian AH, Zeinali S, Abiri M, Sotoudeh S, et al. Genome-wide single nucleotide polymorphism-based autozygosity mapping facilitates identification of mutations in consanguineous families with epidermolysis bullosa. *Exp Derm* 2018b;138:121–31.
- Wierenga KJ, Jiang Z, Yang AC, Mulvihill JJ, Tsinoremas NF. A clinical evaluation tool for SNP arrays, especially for autosomal recessive conditions in offspring of consanguineous parents. *Genet Med* 2013;15:354–60.