bioRxiv preprint doi: https://doi.org/10.1101/530733; this version posted January 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

# Locating suspicious lethal genes by abnormal distributions of SNP patterns

Xiaojun Ding[1],

**1** School of Computer Science and Engineering, Yulin Normal University, Yulin, Guangxi, China

* ding.xiaojun@foxmail.com

## Abstract

A gene, a locatable region of genomic sequence, is the basic functional unit of heredity. Differences in genes lead to the various congenital physical conditions of people. One kind of these major differences are caused by genetic variations named single nucleotide polymorphisms(SNPs). SNPs may affect splice sites, protein structures and so on, and then cause gene abnormities. Some abnormities will lead to fatal diseases. People with these diseases have a small probability of having children. Thus the distributions of SNP patterns on these sites will be different with distributions on other sites. Based on this idea, we present a novel statistical method to detect the abnormal distributions of SNP patterns and then to locate the suspicious lethal genes. We did the test on HapMap data and found 74 suspicious SNPs. Among them, 10 SNPs can map reviewed genes in NCBI database. 5 genes out of them relate to fatal children diseases or embryonic development, 1 gene can cause spermatogenic failure, the other 4 genes are also associated with many genetic diseases. The results validate our idea. The method is very simple and is guaranteed by a statistical test. It is a cheap way to discover the suspicious pathogenic genes and the mutation site. The mined genes deserve further study.

## Author summary

Xiaojun Ding received the BS, MS and PhD degrees in computer science from Central South University. Now he is a assistant professor in Yulin Normal University. His research interests include computational biology and machine learning.

## Introduction

Genes are the most important genetic materials which can determine the health of a person in some ways. The functions of genes may be affected by the genetic variations called SNPs. So it is a good way to study the disease-related genes from SNPs. Many defective genes caused by SNPs for human Mendelian diseases (i.e. single gene diseases) have bAsnaghi2013Ieen found [1, 2]. For example, Prescott el al. [3] found a nonsynonymous SNP in ATG16L1 related to Crohn's disease. Seki et al. [4] reported that a functional SNP in CILP is suspicious to lumbar disc disease. These achievements inspire people.

However, the discovered pathogenic genes caused by SNPs only take a small fraction, most of them are still unknown. At the same time, the number of SNPs is

bioRxiv preprint doi: https://doi.org/10.1101/530733; this version posted January 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

very large and most SNPs do not take effects on genes [5, 6]. Checking all SNPs by biological experiments is a expensive work. Narrowing the range of suspicious SNPs will benefit the study of pathogenic genes greatly [7]. For the purpose, people analyzed SNPs from various angles. Lee et al. [8] builded a functional SNP database which integrates information got from 16 bioinformatics tools and functional SNP effects for disease researches. Cargill et al. [9] studied the different rates of polymorphism within genes and between genes. They concluded that the rates may reflect selection acting against deleterious alleles during evolution and the lower allele frequency of missense cSNPs are possibly associated with diseases. Adzhubei et al. [10] developed a tool named PolyPhen which predicts possible impact of an amino acid substitution on the structure and function of a human protein. Kumar et al. [11] developed a tool named SIFT which predicts whether an amino acid substitution will affect protein function. Their algorithm is suitable to naturally occurring nonsynonymous polymorphisms and laboratory-induced missense mutations. While synonymous mutations can also contribute to human diseases [12]. For example, Westerveld et al. [13] reported that a intronic variants rs1552726 may affect the splice site activity.

In the paper, a novel method is proposed from the angle of genetic law. If a defective gene caused by SNPs can lead to fatal diseases and most of the sick people do not have the change to breed the next generation. This will affect the distributions of the SNPs within the gene. It provides us a novel way to distinguish the pathogenic SNPs from normal SNPs.

## Materials and methods

Given a bi-allele SNP, ‘A‘ and ‘a‘ are used to denote the major and minor allele, respectively. Because chromosomes come in pairs, each individual will take one of the following three SNP patterns: pattern0=‘AA‘, pattern1=‘Aa‘, pattern2=‘aa‘. In a population, the distribution of individuals taking each pattern can be counted. The abnormal distributions are what we concern. Next, an example is given to illustrate the abnormal distributions.

**Fig 1. The heredity of SNPs.**

Example 1: Suppose that there is a distribution for 1000 individuals; 500 individuals out of them take pattern ’AA’ and other people take pattern ’aa’. Nobody takes the pattern ’Aa’.

According to bisexual reproduction rule, a child will inherit one chromosome from his mother and one from his father. If the mother takes pattern ‘AA‘ and the father takes pattern ‘aa‘. The child will take pattern ‘Aa‘. It is shown as Fig 1. If every one has a equal probability to marry other people in the population. The probability of a child with pattern ‘Aa‘ should be 2*0.5*0.5=0.5, that means there should be about 0.5*1000=500 individuals taking the pattern ‘Aa‘. But none is observed. We think that the distribution on this SNP site is abnormal. The reason for the abnormal distribution is probably that the person taking ‘Aa‘ will die in childhood so that we can not observe them. From the analysis, a hypothesis is proposed as following.

**Hypothesis 1:** Some SNP patterns should appear in human populations according to bisexual reproduction but are not observed. These patterns may cause gene abnormities for fatal genetic diseases so that the corresponding person will die in childhood.

In HapMap data, the SNP data of 11 human populations are sequenced. Since the relationships of individuals in each population are unknown, we make a assumption to simplify the computation.

bioRxiv preprint doi: https://doi.org/10.1101/530733; this version posted January 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

**Assumption 1:** In each population, everyone has the equal probability to marry other people and to give a birth to a baby.

For $population_j$, the distribution $P$ of individuals for all the patterns can be counted. $P = [p_0, p_1, p_2]$, where $p_i$ is the percentage of the individuals with $pattern_i$. Under Assumption 1 and the bisexual reproduction rule, the distribution among the next generation (denoted by $P^*$) can be computed according to the distribution $P$. Let $P^* = [p_0^*, p_1^*, p_2^*]$. $P^*$ can be computed by the formulas (1),(2) and (3).

$$p_0{}^* = (p_0 + \frac{p_1}{2})^2 \tag{1}$$

$$p_1{}^* = 2(p_0 + \frac{p_1}{2})(p_2 + \frac{p_1}{2}) \tag{2}$$

$$p_2{}^* = (p_2 + \frac{p_1}{2})^2 \tag{3}$$

If there is no big disaster, the distribution among a human population will not change radically. Under usual circumstances, $P^*$ can be treated as an approximation to the mean distribution of the current population. If $p_i$ is 0, but $p_i^*$ is far from 0. The distribution may be abnormal. Supposing the size of the $population_j$ is $n_j$. The number of individuals matching the $pattern_i$ obeys the binomial distribution. $e_{ij}$ is used to denote the event that the $pattern_i$ is not observed in current $population_j$, the probability of $e_{ij}$ is computed by formula (4)

$$Pr\left(e_{ij}\right) = \left(1 - p_i{}^*\right)^{n_j} \tag{4}$$

In HapMap data, there are 11 human populations. $eAll$ denotes the event that the $pattern_i$ can not be observed in all the populations. The probability of $eAll$ is given by formula (5).

$$p - value\left(eAll\right) = \prod_{j=0}^{10} Pr\left(e_{ij}\right) \tag{5}$$

If the $p - value(eAll)$ is very small, the event that the $pattern_i$ are not observed in all the populations is unlikely. But it actually happens in the observation of HapMap project. The reason may be that the $pattern_i$ may cause gene abnormities for fatal genetic diseases and most patients will die in childhood. In usual statistical test, 0.05 is often used as the threshold of significance level. In our test, since there are many SNPs to be checked. For example, there are 117068 SNPs on chromosome 1. The significance level should be corrected. Here, Bonferroni correction is choosed to adjust the threshold. Given $k$ SNPs, $3k$ hypothesizes need to be tested. The p-value threshold of significance should be $0.05/3k$. The SNPs with $p - value(eAll)$ below the p-value threshold are suspicious. After finding these SNPs, the suspicious lethal genes can be located by NCBI database.

The HapMap [14] data (genotypes data of phase 3.3 consensus) is used to validate our method. The HapMap data contains 11 human populations. 1417 individuals are sequenced. The detail of human populations is listed in Table 1.

# Results

In our experiment,the statistical test is checked on each chromosome separately. For one chromosome, the common bi-allele SNPs in the 11 populations are extracted. On chromosome 1, there are 117068 common SNPs and 8 SNPs are detected to be

bioRxiv preprint doi: https://doi.org/10.1101/530733; this version posted January 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

**Table 1. information of human populations.**

| Label | Population sample | Size |
|---|---|---|
| ASW | African ancestry in Southwest USA | 87 |
| CEU | Utah residents with Northern and Western European ancestry from the CEPH collection | 174 |
| CHB | Han Chinese in Beijing, China | 139 |
| CHD | Chinese in Metropolitan Denver, Colorado | 109 |
| GIH | Gujarati Indians in Houston, Texas | 101 |
| JPT | Japanese in Tokyo, Japan | 116 |
| LWK | Luhya in Webuye, Kenya | 110 |
| MEX | Mexican ancestry in Los Angeles, California | 86 |
| MKK | Maasai in Kinyawa, Kenya | 184 |
| TSI | Toscans in Italy | 102 |
| YRI | Yoruba in Ibadan, Nigeria | 209 |

suspicious. On the whole 22 chromosomes, 74 SNPs are obtained among 1,395,560 SNPs (we do not check SNPs on the chromosome X and Y because the heredity on chromosome X and Y are different with that on autosomes). By looking up in NCBI database, 29 genes are located. Some of them have been studied in many years and some are still new to people. According to the reference sequence (RefSeq) status code, these genes can be divided into 3 catalogs.

A reviewed gene means that its RefSeq record has been reviewed by NCBI staff or by a collaborator. The NCBI review process includes assessing available sequence data and the literature. Some RefSeq records may incorporate expanded sequence and annotation information.

In our results, 10 SNPs map reviewed genes. The corresponding chromosomes, SNPs, genes, gene types, alleles, disease patterns and p-values are listed in Table 2. For the first two SNPs and their patterns, the expectations of the number of individuals in each population are listed in Table 3 and Table 4, respectively. For the pattern 'AA' at SNP site rs2145402, there should be about 34.4 individuals in all populations. But none is observed. If the pattern 'AA' is normal, the probability of the event is only 3.05E-16. It is too small to happen by chance. So we think the distribution of the pattern at SNP rs2145402 is abnormal. The corresponding gene LYST is suspicious to genetic diseases.

**Table 2. information of SNPs mapping reviewed genes.**

| chr | SNP | gene | alleles | disease pattern | pvalue |
|---|---|---|---|---|---|
| 1 | rs2145402 | LYST | A/C | AA | 3.05E-16 |
| 1 | rs4915931 | ROR1 | A/G | AA | 1.70E-12 |
| 1 | rs4660992 | BMP8B | C/T | TT | 9.91E-10 |
| 6 | rs9263745 | CCHCR1 | A/G | AA | 5.67E-11 |
| 7 | rs11766679 | DPP6 | A/G | GG | 4.29E-10 |
| 10 | rs12263497 | INPP5F | A/G | GG | 1.46E-10 |
| 11 | rs1552726 | NLRP14 | A/G | GG | 2.64E-09 |
| 14 | rs3742943 | JAG2 | C/T | TT | 6.07E-08 |
| 16 | rs1646233 | CBFA2T3 | A/G | AA | 1.20E-09 |
| 22 | rs11705619 | TXNRD2 | C/T | CC | 1.43E-09 |

The reviewed genes are familiar to people. There are relatively more studies on them. The number of literatures focused on this kind of genes are more than that of

bioRxiv preprint doi: https://doi.org/10.1101/530733; this version posted January 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

**Table 3. expectations of the number of individuals in each population who takes pattern 'AA' at SNP site rs2145402.**

| **population** | CHB | CHD | JPT | CEU | TSI | ASW | LWK | MKK | YRI | GIH | MEX | All |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **expectation** | 1.1 | 0.4 | 0.5 | 11.6 | 8.2 | 1.0 | 0.0 | 0.1 | 0.1 | 4.8 | 6.6 | 34.4 |

**Table 4. expectations of the number of individuals in each population who takes pattern 'AA' at SNP site rs4915931.**

| **population** | CHB | CHD | JPT | CEU | TSI | ASW | LWK | MKK | YRI | GIH | MEX | All |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **expectation** | 0.2 | 0.1 | 0.2 | 8.8 | 4.2 | 2.5 | 1.3 | 0.3 | 2.5 | 4.1 | 2.5 | 26.7 |

other genes. These related literatures provide us a intuitive impression and are listed in the following. From the details we can see that most of these genes are actually associated with fatal genetic diseases.

1) SNP rs2145402 maps gene LYST.

In ClinVar database [15], LYST is associated with lung cancer, Malignant melanoma and Chediak-Higashi syndrome. Many researchers [16–21] also reported that gene LYST is associated with Chediak-Higashi syndrome. Chediak-Higashi syndrome can affect many parts of the body, particularly the immune system. The disease damages immune system cells. Most affected individuals have repeated and persistent infections starting in infancy or early childhood [22]. The result of the disease is very serious and most affected individuals die in childhood [23].

2) SNP rs4915931 maps gene ROR1

In ClinVar database [15], ROR1 is associated with malignant melanoma. Broome et al. [24] reported that ROR1 is a receptor tyrosine kinase expressed during embryogenesis, on chronic lymphocytic leukemia and in other malignancies. Hudecek et al. [25] found that ROR1 is highly expressed during early embryonic development but expressed at very low levels in adult tissues. Many papers reported that ROR1 has a very close relation with chronic lymphocytic leukemia [26, 27] and acute lymphoblastic leukemia [28–32]. ROR1 is suggested as the targeted therapy for human malignancies [33, 34].

3) SNP rs4660992 maps gene BMP8B

BMP8B is a thermogenic protein which increases brown adipose tissue thermogenesis through both central and peripheral actions and regulates energy balance in partnership with hypothalamic AMPK [35]. Zhao et al. [36] showed that mouse Bmp8a (Op2) and Bmp8b play a role in spermatogenesis and placental development. Ying et al. [37] reported that BMP8B is required for the generation of primordial germ cells in the mouse.

4) SNP rs11766679 maps gene DPP6

Genetic variation in DPP6 is associated with amyotrophic lateral sclerosis [20, 38, 39] and Familial Idiopathic Ventricular Fibrillation [40]. Golz et al. [41] found that DPP6 is suspicious to many diseases such as cardiovascular diseases, endocrinological diseases, metabolic diseases, gastroenterological diseases, cancer, hematological diseases, inflammation, muscle skeleton diseases, neurological diseases, urological diseases, reproduction disorders and respiratory diseases.

5) SNP rs12263497 maps gene Inpp5f

Zhu el al. [42] reported that Inpp5f is a polyphosphoinositide phosphatase that regulates cardiac hypertrophic responsiveness. Kim et al. [43] found that INPP5F inhibits STAT3 activity and suppresses gliomas tumorigenicity. Palermo et al. [44] reported that gene expression of INPP5F can be as an independent prognostic marker in fludarabine-based therapy of chronic lymphocytic leukemia. Bai et al. [45] reported that alteration of Akt signal plays an important role in diabetic cardiomyopathy.

bioRxiv preprint doi: https://doi.org/10.1101/530733; this version posted January 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Inpp5f is a negative regulator of Akt signaling.

6) SNP rs9263745 maps gene CCHCR1

CCHCR1 is associated with malignant melanoma in ClinVar database [15]. CCHCR1 is up-regulated in skin cancer and associated with EGFR expression [46]. The CCHCR1 (HCR) gene is relevant for skin steroidogenesis and downregulated in cultured psoriatic keratinocytes [47].

7) SNP rs1552726 maps gene NLRP14

NLRP14 may play a regulatory role in the innate immune system [48]. Mutations in the testis-specific NALP14 gene in men suffering from spermatogenic failure in GeneCards database [49]. Westerveld et al. [13] collected the data of 157 patients, they identified 25 suspicious variants in total: 1 nonsense mutation, 14 missense mutations, 6 silent mutations and 4 intronic variants. By using ESEfinder and SpliceSiteFinder to check these SNPs, only the SNP rs1552726 is predicted to affect the correct splicing. Abe et al. [50] reported that germ-cell-specific inflammasome component NLRP14 negatively regulates cytosolic nucleic acid sensing to promote fertilization.

8) SNP rs3742943 maps gene JAG2

GO annotations related to JAG2 include Notch binding and calcium ion binding. The gene serve as a ligand for Notch signaling receptors. The Notch signaling pathway is an intercellular signaling mechanism that is essential for proper embryonic development. Defect in JAG2 may cause ossifying fibroma and shipyard eye in GeneCards database [49]. Houde et al. [51] observed the overexpression of the NOTCH ligand JAG2 in malignant plasma cells from multiple myeloma patients and cell lines. Yustein et al. [52] validated that induction of ectopic Myc target gene JAG2 augments hypoxic growth and tumorigenesis in a human B-cell model. Asnaghi et al. [53] reported that a role for Jag2 promotes uveal melanoma dissemination and growth. Vaish et al. [54] reported that JAG2 enhances tumorigenicity and chemoresistance of colorectal cancer cells.

9) SNP rs1646233 maps gene CBFA2T3

CBFA2T3-GLIS2 Fusion Protein Defines an Aggressive Subtype of Pediatric Acute Megakaryoblastic Leukemia [55] and CBFA2T3-GLIS2 fusion transcript is a novel common feature in pediatric, cytogenetically normal AML, not restricted to FAB M7 subtype [56]. CBFA2T3-GLIS2-positive is closed related to pediatric acute megakaryoblastic leukemia with non-Down syndrome [57].

10) SNP rs11705619 maps gene TXNRD2

Mutations in the gene TXNRD2 cause dilated cardiomyopathy [58] Jakupoglu et al. [59] did the experiment of Txnrd2 deletion and found which leads to fatal dilated cardiomyopathy and morphological abnormalities of cardiomyocytes. Prasad et al. [60] reported that TXNRD2 knockout is embryonic lethal in mice due to cardiac malformation.

## Conclusion

In the paper, we gain enlightenment from Mendel's genetic experiments and propose a simple method which can utilize the distributions of SNP patterns among human populations to mine the pathogenic genes. On HapMap data, 74 SNPs are selected in 22 autosomal chromosomes, 10 SNPs can map reviewed genes in NCBI database. Among these genes, LYST gene and ROR1 gene are reported to relate to fatal genetic children diseases. Genes JAG2, TXNRD2 and BMP8B play important roles in embryonic development and lead to many fatal diseases. NALP14 gene may cause spermatogenic failure. Among 25 suspicious variants, only SNP rs1552726 is predicted to affect the correct splicing of gene NLRP14 [13], rs1552726 is also one of the 10 SNPs which maps reviewed genes. The left genes DPP6,Inpp5f,CCHCR1 and

bioRxiv preprint doi: https://doi.org/10.1101/530733; this version posted January 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

CBFA2T3 are also associated with many genetic diseases. Looked from the overall, we think the results are good and can validate our idea in some ways. The method can give a narrow range of suspicious pathogenic genes which deserve further studies. As whole-genome sequencing advances, more and more data can be achieved, the method can get more accurate and interesting results. The method is a simple and cheap way to find the suspicious pathogenic genes and SNPs.

## Acknowledgments

## References

1. Sellick GS, Longman C, Tolmie J, Newbury Ecob R, Geenhalgh L, Hughes S, et al. Genomewide linkage searches for Mendelian disease loci can be efficiently conducted using high density SNP genotyping arrays. Nucleic Acids Res. 2004;32(20):e164.

2. Pico AR, Smirnov IV, Chang JS, Yeh RF, Wiemels JL, Wiencke JK, et al. SNPLogic: an interactive single nucleotide polymorphism selection, annotation, and prioritization system. Nucleic acids research. 2009;37(suppl 1):D803 D809.

3. Prescott NJ, Fisher SA, Franke A, Hampe J, Onnie CM, Soars D, et al. A nonsynonymous SNP in ATG16L1 predisposes to ileal Crohn's disease and is independent of CARD15 and IBD5. Gastroenterology. 2007;132(5):1665 1671.

4. Seki S, Kawaguchi Y, Chiba K, Mikami Y, Kizawa H, Oya T, et al. A functional SNP in CILP, encoding cartilage intermediate layer protein, is associated with susceptibility to lumbar disc disease. Nature genetics. 2005;37(6):607 612.

5. Ding X, Wang J, Zelikovsky A, Guo X, Xie M, Pan Y. Searching High-Order SNP Combinations for Complex Diseases Based on Energy Distribution Difference. Computational Biology and Bioinformatics, IEEE/ACM Transactions on. 2015;12(3):695–704. doi:10.1109/TCBB.2014.2363459.

6. Cai Z, Sabaa H, Wang Y, Goebel R, Wang Z, Xu J, et al. Most parsimonious haplotype allele sharing determination. BMC bioinformatics. 2009;10(1):115.

7. Guo X, Meng Y, Yu N, Pan Y. Cloud computing for detecting high order genome wide epistatic interaction via dynamic clustering. BMC Bioinformatics. 2014;15(1).

8. Lee PH, Shatkay H. F SNP: computationally predicted functional SNPs for disease association studies. Nucleic acids research. 2008;36(suppl 1):D820 D824.

9. Cargill M, Altshuler D, Ireland J, Sklar P, Ardlie K, Patil N, et al. Characterization of single nucleotide polymorphisms in coding regions of human genes. Nature genetics. 1999;22(3):231 238.

10. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. Nature methods. 2010;7(4):248 249.

11. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non synonymous variants on protein function using the SIFT algorithm. Nature protocols. 2009;4(7):1073 1081.

bioRxiv preprint doi: https://doi.org/10.1101/530733; this version posted January 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

12. Sauna ZE, Kimchi Sarfaty C. Understanding the contribution of synonymous mutations to human disease. Nature Reviews Genetics. 2011;12(10):683 691.

13. Westerveld G, Korver C, van Pelt A, Leschot N, Van der Veen F, Repping S, et al. Mutations in the testis specific NALP14 gene in men suffering from spermatogenic failure. Human Reproduction. 2006;21(12):3178 3184.

14. NCBI. type [; 2017]Available from: `https://www.ncbi.nlm.nih.gov/probe/docs/projhapmap/`.

15. NCBI. type [; 2017]Available from: `http://www.ncbi.nlm.nih.gov/clinvar`.

16. Al Tamemi S, Al Zadjali S, Al Ghafri F, Dennison D. Chediak Higashi Syndrome: Novel Mutation of the CHS1/LYST Gene in 3 Omani Patients. Journal of pediatric hematology/oncology. 2014;36(4):e248 e250.

17. Weisfeld Adams JD, Mehta L, Rucker JC, Dembitzer FR, Szporn A, Lublin FD, et al. A typical Chediak Higashi syndrome with attenuated phenotype: three adult siblings homozygous for a novel LYST deletion and with neurodegenerative disease. Orphanet journal of rare diseases. 2013;8(1):46.

18. Anistoroaei R, Krogh A, Christensen K. A frameshift mutation in the LYST gene is responsible for the Aleutian color and the associated Chediak–Higashi syndrome in American mink. Animal genetics. 2013;44(2):178 183.

19. Kaya Z, Ehl S, Albayrak M, Maul Pavicic A, Schwarz K, Kocak U, et al. A novel single point mutation of the LYST gene in two siblings with different phenotypic features of Chediak Higashi syndrome. Pediatric blood cancer. 2011;56(7):1136 1139.

20. Del Bo R, Ghezzi S, Corti S, Santoro D, Prelle A, Mancuso M, et al. DPP6 gene variability confers increased risk of developing sporadic amyotrophic lateral sclerosis in Italian patients. Journal of Neurology, Neurosurgery Psychiatry. 2008;79(9):1085 1085.

21. Zarzour W, Kleta R, Frangoul H, Suwannarat P, Jeong A, Kim SY, et al. Two novel CHS1 ( LYST ) mutations: Clinical correlations in an infant with Chediak higashi syndrome. Molecular genetics and metabolism. 2005;85(2):125 132.

22. GeneticsHomeReference. type [; 2017]Available from: `http://ghr.nlm.nih.gov/condition/chediak-higashi-syndrome`.

23. HealthGradesInc. type [; 2017]Available from: `http://www.rightdiagnosis.com/c/chediakhigashisyndrome/deaths.htm`.

24. Broome HE, Rassenti LZ, Wang HY, Meyer LM, Kipps TJ. ROR1 is expressed on hematogones (non neoplastic human B lymphocyte precursors) and a minority of precursor B acute lymphoblastic leukemia. Leukemia research. 2011;35(10):1390 1394.

25. Hudecek M, Schmitt TM, Baskar S, Lupo Stanghellini MT, Nishida T, Yamamoto TN, et al. The B cell tumor associated antigen ROR1 can be targeted with T cells modified to express a ROR1 specific chimeric antigen receptor. Blood. 2010;116(22):4532 4541.

26. DaneshManesh AH, Mikaelsson E, Jeddi Tehrani M, Bayat AA, Ghods R, Ostadkarampour M, et al. Ror1, a cell surface receptor tyrosine kinase is expressed in chronic lymphocytic leukemia and may serve as a putative target for therapy. International journal of cancer. 2008;123(5):1190 1195.

bioRxiv preprint doi: https://doi.org/10.1101/530733; this version posted January 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

27. Hojjat Farsangi M, Khan AS, Daneshmanesh AH, Moshfegh A, Sandin s, Mansouri L, et al. The Tyrosine Kinase Receptor ROR1 Is Constitutively Phosphorylated in Chronic Lymphocytic Leukemia (CLL) Cells. PLoS ONE. 2013;8(10):e78339. doi:10.1371/journal.pone.0078339.

28. Shaheen I, Ibrahim N. Detection of Orphan Receptor Tyrosine Kinase (ROR 1) Expression in Egyptian Pediatric Acute Lymphoblastic Leukemia. Fetal Pediatric Pathology. 2012;31(3):113 119.

29. Shabani M, Asgarian Omran H, Vossough P, Sharifian RA, Faranoush M, Ghragozlou S, et al. Expression profile of orphan receptor tyrosine kinase (ROR1) and Wilms' tumor gene 1 (WT1) in different subsets of B cell acute lymphoblastic leukemia. Leukemia lymphoma. 2008;49(7):1360 1367.

30. Milani L, Lundmark A, Kiialainen A, Nordlund J, Flaegstad T, Forestier E, et al. DNA methylation for subtype classification and prediction of treatment outcome in patients with childhood acute lymphoblastic leukemia. Blood. 2010;115(6):1214 1225.

31. Bicocca VT, Chang BH, Masouleh BK, Muschen M, Loriaux MM, Druker BJ, et al. Crosstalk between ROR1 and the Pre B cell receptor promotes survival of t (1; 19) acute lymphoblastic leukemia. Cancer cell. 2012;22(5):656 667.

32. Dave H, Anver MR, Butcher DO, Brown P, Khan J, Wayne AS, et al. Restricted cell surface expression of receptor tyrosine kinase ROR1 in pediatric B lineage acute lymphoblastic leukemia suggests targetability with therapeutic monoclonal antibodies. PloS one. 2012;7(12):e52655.

33. Baskar S, Wiestner A, Wilson WH, Pastan I, Rader C. Targeting malignant B cells with an immunotoxin against ROR1. In: MAbs. vol. 4. Taylor & Francis; 2012. p. 349–361.

34. Guilly Rebagay SY, Liu C, Cheung NK. ROR1 and ROR2 in human malignancies: potentials for targeted therapy. Frontiers in oncology. 2012;2.

35. Whittle AJ, Carobbio S, Martins L, Slawik M, Hondares E, Vázquez MJ, et al. BMP8B increases brown adipose tissue thermogenesis through both central and peripheral actions. Cell. 2012;149(4):871 885.

36. Zhao GQ, Hogan BL. Evidence that mouse Bmp8a (Op2) and Bmp8b are duplicated genes that play a role in spermatogenesis and placental development. Mechanisms of development. 1996;57(2):159 168.

37. Ying Y, Liu XM, Marble A, Lawson KA, Zhao GQ. Requirement of Bmp8b for the generation of primordial germ cells in the mouse. Molecular Endocrinology. 2000;14(7):1053 1063.

38. van Es MA, van Vught PW, Blauw HM, Franke L, Saris CG, Van Den Bosch L, et al. Genetic variation in DPP6 is associated with susceptibility to amyotrophic lateral sclerosis. Nature genetics. 2007;40(1):29 31.

39. Daoud H, Valdmanis PN, Dion PA, Rouleau GA. Analysis of DPP6 and FGGY as candidate genes for amyotrophic lateral sclerosis. Amyotrophic Lateral Sclerosis. 2010;11(4):389 391.

bioRxiv preprint doi: https://doi.org/10.1101/530733; this version posted January 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

40. Alders M, Koopmann TT, Christiaans I, Postema PG, Beekman L, Tanck MW, et al. Haplotype Sharing Analysis Implicates Chromosome 7q36 Harboring DPP6 in Familial Idiopathic Ventricular Fibrillation. The American Journal of Human Genetics. 2009;84(4):468 476.

41. Golz S, Bruggemeir U, Geerts A, Schubel A, Zubov D. type [; 2005].

42. Zhu W, Trivedi CM, Zhou D, Yuan L, Lu MM, Epstein JA. Inpp5f is a polyphosphoinositide phosphatase that regulates cardiac hypertrophic responsiveness. Circulation research. 2009;105(12):1240 1247.

43. Kim HS, Li A, Ahn S, Song H, Zhang W. Inositol Polyphosphate 5 Phosphatase F (INPP5F) inhibits STAT3 activity and suppresses gliomas tumorigenicity. Scientific reports. 2014;4.

44. Palermo G, Maisel D, Barrett M, Smith H, Duchateau-Nguyen G, Nguyen T, et al. Gene expression of INPP5F as an independent prognostic marker in fludarabine-based therapy of chronic lymphocytic leukemia. Blood cancer journal. 2015;5(10):e353.

45. Bai D, Zhang Y, Shen M, Sun Y, Xia Q, Zhang Y, et al. Hyperglycemia and hyperlipidemia blunts the Insulin-Inpp5f negative feedback loop in the diabetic heart. Scientific reports. 2016;6:22068.

46. Suomela S, Elomaa O, Skoog T, Ala aho R, Jeskanen L, Prssinen J, et al. CCHCR1 is up regulated in skin cancer and associated with EGFR expression. PloS one. 2009;4(6):e6030.

47. Tiala I, Suomela S, Huuhtanen J, Wakkinen J, Hltt Vuori M, Kainu K, et al. The CCHCR1 (HCR) gene is relevant for skin steroidogenesis and downregulated in cultured psoriatic keratinocytes. Journal of Molecular Medicine. 2007;85(6):589 601.

48. NCBI. type [; 2017]Available from: `http://www.ncbi.nlm.nih.gov/gene/338323`.

49. GENECARDS. type [; 2017]Available from: `http://www.genecards.org/`.

50. Abe T, Lee A, Sitharam R, Kesner J, Rabadan R, Shapira SD. Germ-Cell-Specific inflammasome component NLRP14 negatively regulates cytosolic nucleic acid sensing to promote fertilization. Immunity. 2017;46(4):621–634.

51. Houde C, Li Y, Song L, Barton K, Zhang Q, Godwin J, et al. Overexpression of the NOTCH ligand JAG2 in malignant plasma cells from multiple myeloma patients and cell lines. Blood. 2004;104(12):3697 3704.

52. Yustein JT, Liu YC, Gao P, Jie C, Le A, Vuica Ross M, et al. Induction of ectopic Myc target gene JAG2 augments hypoxic growth and tumorigenesis in a human B cell model. Proceedings of the National Academy of Sciences. 2010;107(8):3534 3539.

53. Asnaghi L, Handa JT, Merbs SL, Harbour JW, Eberhart CG. A role for Jag2 in promoting uveal melanoma dissemination and growth. Investigative ophthalmology visual science. 2013;54(1):295 306.

54. Vaish V, Kim J, Shim M. Jagged 2 (JAG2) enhances tumorigenicity and chemoresistance of colorectal cancer cells. Oncotarget. 2017;8(32):53262.

bioRxiv preprint doi: https://doi.org/10.1101/530733; this version posted January 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

55. Gruber TA, Larson Gedman A, Zhang J, Koss CS, Marada S, Ta HQ, et al. An Inv (16)(p13. 3q24. 3) Encoded CBFA2T3 GLIS2 Fusion Protein Defines an Aggressive Subtype of Pediatric Acute Megakaryoblastic Leukemia. Cancer cell. 2012;22(5):683 697.

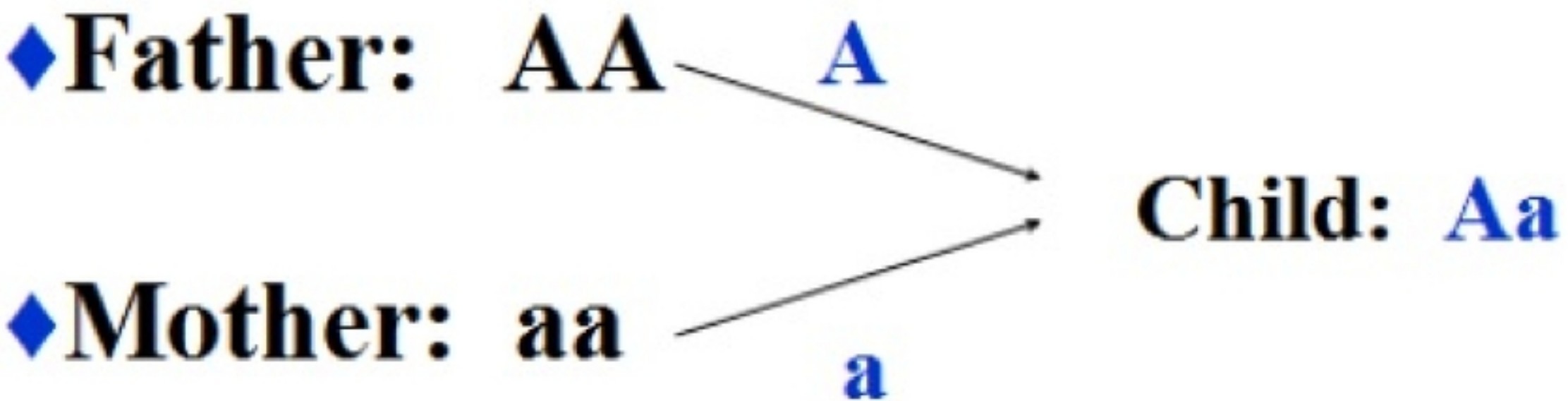
56. Masetti R, Pigazzi M, Togni M, Astolfi A, Indio V, Manara E, et al. CBFA2T3 GLIS2 fusion transcript is a novel common feature in pediatric, cytogenetically normal AML, not restricted to FAB M7 subtype. Blood. 2013;121(17):3469 3472.

57. Hara Y, Yamato G, Shiba N, Ohki K, Park MJ, Tomizawa D, et al.. High BMP2 Expression Is a Poor Prognostic Factor and a Good Candidate to Identify CBFA2T3-GLIS2-like High-Risk Subgroup in Pediatric Acute Myeloid Leukemia; 2015.

58. Sibbing D, Pfeufer A, Perisic T, Mannes AM, Fritz Wolf K, Unwin S, et al. Mutations in the mitochondrial thioredoxin reductase gene TXNRD2 cause dilated cardiomyopathy. European heart journal. 2011;32(9):1121 1133.

59. Jakupoglu C, Przemeck GK, Schneider M, Moreno SG, Mayr N, Hatzopoulos AK, et al. Cytoplasmic thioredoxin reductase is essential for embryogenesis but dispensable for cardiac development. Molecular and cellular biology. 2005;25(5):1980 1988.

60. Prasad R, Hughes C, Chan L, Peters C, Nathwani N, Clark A, et al. A mutation in thioredoxin reductase 2 (TXNRD2) is associated with a predominantly adrenal phenotype in humans. 2013;.




Figure