Gene mapping and molecular analysis of hereditary non-polyposis colorectal cancer (Lynch Syndrome) using systems biological approaches

Mahmood Rasool¹*, Sajjad Karim¹, Muhammad Imran Naseer¹, Peter Natesan Pushparaj¹, Adel Abuzenadah¹,², Mohammed Hussein Al-Qahtani¹

¹Center of Excellence in Genomic Medicine Research, King Abdulaziz University, Jeddah, Saudi Arabia; ²Department of Medical Laboratory Technology, Faculty of Applied Medical Sciences, King Abdulaziz University, Jeddah, Saudi Arabia. Dr. Mahmood Rasool - Email: mahmoodrasool@yahoo.com; *Corresponding Author

Received January 26, 2019; Accepted March 1, 2019; Published April 15, 2019

DOI: 10.6026/97320630015269

Abstract:
Hereditary non-polyposis colorectal cancer (HNPCC) also known as Lynch Syndrome (LS), is a hereditary form of colorectal cancer (CRC). LS is caused by mutations in the mismatch repair (MMR) genes, mostly in MLH1, MSH2, MSH6 and PMS2. Identification of these gene mutations is essential to diagnose CRC, especially at a young age to increase the survival rate. Using open target platform, we have performed genetic association studies to analyze the different genes involved in the LS and to obtain target evidence. We have also analyzed upstream regulators as target molecules in the data sets. We discovered that MLH1, MSH2, MSH6, PMS2, MLH3, EPCAM, TGFBR2, FBXO11 and PRSS58 were showing most association in LS. Our findings may further enhance the understanding of the hereditary form of CRC.

Keywords: Lynch syndrome, HNPCC, Mismatch repair genes, Open target platform

Background:
Lynch syndrome is an autosomal dominant condition caused by many mismatch repair genes including four important genes; MLH1, PMS2, MSH2 and MSH6 [1]. LS was named in honor of Henry T. Lynch, who reported several families in detail during 1966-67 [2, 3]. LS accounts for 1-5% of all CRC and also presents an increased risk of many extra colonic cancer types [4]. Mutations in the MMR genes lead to the inactivation or lower efficiency to repair mismatches in DNA that leads to the accumulation of spontaneous mutations mostly consist of the insertions and deletions in short repetitive DNA sequences termed microsatellites. The changes in short microsatellite sequences lead to the microsatellite instability, that is found in the majority of LS tumors (>90%) in patients with germ line mutations in MMR genes [5]. Therefore the current strategy before sequencing these MMR genes is to do the microsatellite instability (MSI) testing. So, if the patient tumor DNA is found with MSI, it will likely yield a mutation in MMR genes. Generally, the five different regions with microsatellites are looked at, and the tumor is considered highly unstable if instability is found in two or more regions. While the tumor is called as unstable-low if the instability is found in only one region and stable if no instability is found [6-9]. The identification of the MMR gene status is very important for surveillance and early intervention especially in the carriers and the family members of the CRC patients, therefore appropriate measure could be taken to limit the disease and improve the survival of the patients and carriers of the disease. And also excluding the family members for any mismatch
gene mutation carriers may reduce the worry and high-risk surveillance burden of prevention testing. In this article, we have utilized open target platform for genetic association studies of different genes spectrum involved in LS, their upstream regulators and canonical pathways.

Materials & Methods:
Data-Mining of Genetic Associations in Lynch Syndrome:
We have used the Open Targets Platform (https://www.targetvalidation.org/), a free online integrated web resource of genetics, omics and chemical data to aid systematic drug target identification and ranking linking these associations back to the underlying evidence and its source which gives the prioritization of drugs for gene targets based on the strength of their association with a particular disease such as LS [10, 11]. The open targets platform assemble data types from multiple open sources and implement a scoring system on the gene target-disease associations aiming at providing users to classify, recognize, and prioritize suitable drug targets for further examination. The Open Targets score for the associations is a range between 0 (no association) and 1 (strong association). The Open Targets Platform gives scores with varying shades of blue (the darker the blue, the stronger the genetic association with a particular disease) and the overall association score is the result of the combination of all data source scores [10, 11].

### Table 1: Genetic associations observed in Lynch Syndrome

| Gene Symbol | Gene Target ID | Overall Score | Genetic Association Score | Somatic Mutation Score | Gene Name |
|-------------|----------------|---------------|----------------------------|------------------------|-----------|
| MLH1        | ENSG00000076242 | 1             | 0                          | 0                      | mutL homolog 1 |
| MSH2        | ENSG00000095002 | 1             | 0                          | 0                      | mutS homolog 2 |
| MSH6        | ENSG00000160621 | 1             | 0                          | 0                      | mutS homolog 6 |
| PMS2        | ENSG00000122512 | 1             | 0                          | 0                      | PMS1 homolog 2, mismatch repair system component |
| EPCAM       | ENSG00000198881 | 1             | 0                          | 0                      | epithelial cell adhesion molecule |
| TGFBR2      | ENSG00000163513 | 1             | 0.694444445                | 0                      | transforming growth factor beta receptor 2 |
| MLH3        | ENSG00000190841 | 1             | 0                          | 0                      | mutL homolog 3 |
| FBXO11      | ENSG00000380910 | 0.834652      | 0                          | 0                      | F-box protein 11 |
| PRSS50      | ENSG00000025022 | 0.498263889   | 0                          | 0                      | serine protease 58 |

### Table 2: Upstream regulators and the corresponding target genes observed in Lynch syndrome

| S. No | Upstream Regulator | Molecule Type | Gene Target ID | Target Genes |
|-------|--------------------|---------------|---------------|--------------|
| 1     | MBD4               | enzyme        | C2C8-71       | MSH1, MSH2, MSH6, PMS2 |
| 2     | PTGS-1             | transcription regulator | 2.11E-12 | MSH1, MSH2, MSH6, PMS2, Vegf, VEGFA |
| 3     | CDH11              | enzyme        | 0.00000029    | TGFBR2, Vegf, VEGFA |
| 4     | TP53               | transcription regulator | 0.00000096 | MSH1, MSH2, MSH6, PMS2, TGFBR2, TYMS, Vegf, VEGFA |
| 5     | MYC                | transcription regulator | 0.00000092 | CD274, LEPCAM, MSH2, TGFBR2, TYMS, Vegf, VEGFA |
| 6     | PBLN2              | other         | 0.00001582    | Vegf, VEGFA |
| 7     | CCNA1              | other         | 0.0000182     | Vegf, VEGFA |
| 8     | IL18               | cytokine      | 0.0000225     | MSH2, TGFBR2, Vegf, VEGFA |
| 9     | MSH6               | other         | 0.0000303     | Vegf, VEGFA |
| 10    | PDGF-D             | growth factor | 0.0000454    | Vegf, VEGFA |
| 11    | SPI2               | transcription regulator | 0.0000533   | MSH6, TGFBR2, Vegf, VEGFA |
| 12    | LEPCAM             | transcription regulator | 0.0000539 | MSH1, MSH2, MSH6, Vegf, VEGFA |
| 13    | EGLN1              | enzyme        | 0.0000599     | TGFBR2, Vegf, VEGFA |
| 14    | Kras               | enzyme        | 0.0000666     | CD274, LEPCAM, TGFBR2, Vegf, VEGFA |
| 15    | F3                 | Trans membrane receptor | 0.000072  | TGFBR2, Vegf, VEGFA |
| 16    | Lpl2               | group         | 0.0000846     | Vegf, VEGFA |
| 17    | IL7/A              | cytokine      | 0.0000885     | CD274, TGFBR2, Vegf, VEGFA |
| 18    | SPI2               | transcription regulator | 0.000101 | MSH6, TGFBR2, TYMS, Vegf, VEGFA |
| 19    | MLH1               | enzyme        | 0.0001099     | MSH1, MSH2 |
| 20    | HRBBP              | other         | 0.0001099     | Vegf, VEGFA |
| 21    | LDL2               | complex       | 0.0001292     | MSH1, TGFBR2, Vegf, VEGFA |
| 22    | CYP4F2             | enzyme        | 0.0001386     | Vegf, VEGFA |
| 23    | RhoA               | transcription regulator | 0.000156 | Vegf, VEGFA |
| 24    | S100A7             | other         | 0.000166      | Vegf, VEGFA |
| 25    | NFKB (complex)     | complex       | 0.000184      | CD274, LEPCAM, MSH2, Vegf, VEGFA |
| 26    | mir-150-5p (and other miRNAs w/seed CUC.CCA) | mature micro RNA | 0.000199 | Vegf, VEGFA |
| 27    | SERPINE3          | other         | 0.000205      | Vegf, VEGFA |
| 28    | HHX                | transcription regulator | 0.000275 | Vegf, VEGFA |
| 29    | mir-155            | Micro RNA     | 0.000289      | MSH1, MSH2 |
| 30    | MRT1               | kinase        | 0.000314      | CD274, Vegf, VEGFA |
| 31    | NRP1               | other         | 0.000337      | TGFBR2, VEGFA |
| No. | Gene Symbol | Category | Gene Name | p-Value |
|-----|-------------|----------|-----------|---------|
| 32  | RLN2        | Other    |           | 3.1e-03  |
| 33  | FGFR1       | Group    |           | 3.2e-03  |
| 34  | CBS4        | Enzyme   | Fatty Acid | 4.4e-03  |
| 35  | AOX1        | Enzyme   |           | 4.5e-03  |
| 37  | PAPD6       | Other    |           | 4.6e-03  |
| 38  | Angio tensin II receptor type 1 | Group |           | 6.3e-03  |
| 39  | HMOX1       | Enzyme   |           | 6.4e-03  |
| 40  | TP75        | Transcription regulator |           | 6.5e-03  |
| 41  | FOXM1       | Transcription regulator |           | 6.6e-03  |
| 42  | PI3K (family) | Group    |           | 6.7e-03  |
| 43  | VHL         | Transcription regulator |           | 6.8e-03  |
| 44  | SMAD7       | Transcription regulator |           | 6.9e-03  |
| 45  | Histone h4 | Group    |           | 7.0e-03  |
| 46  | HTATIP2     | Transcription regulator |           | 7.1e-03  |
| 47  | RELA        | Transcription regulator |           | 7.2e-03  |
| 48  | IL15        | Cytokine  |           | 7.3e-03  |
| 49  | IL4         | Cytokine  |           | 7.4e-03  |
| 50  | APEX1       | Enzyme   |           | 7.5e-03  |
| 52  | ALB         | Transporter |           | 7.7e-03  |
| 53  | Angiopoietin  | Group |           | 7.8e-03  |
| 54  | ALDOC       | Enzyme   |           | 7.9e-03  |
| 55  | TNFR1       | Transporter |           | 8.0e-03  |
| 56  | IL12        | Cytokine  |           | 8.1e-03  |
| 57  | TNFR2       | Transporter |           | 8.2e-03  |
| 58  | APOB        | Transporter |           | 8.3e-03  |
| 59  | Angiopoietin  | Group |           | 8.4e-03  |
| 60  | ALDH1       | Transporter |           | 8.5e-03  |
| 61  | IL13        | Cytokine  |           | 8.6e-03  |
| 62  | TNFR2       | Transporter |           | 8.7e-03  |
| 63  | IL12        | Cytokine  |           | 8.8e-03  |
| 64  | APOB        | Transporter |           | 8.9e-03  |
| 65  | Angiopoietin  | Group |           | 9.0e-03  |
| 66  | ALDH1       | Transporter |           | 9.1e-03  |
| 67  | IL13        | Cytokine  |           | 9.2e-03  |
| 68  | TNFR2       | Transporter |           | 9.3e-03  |
| 69  | IL12        | Cytokine  |           | 9.4e-03  |
| 70  | APOB        | Transporter |           | 9.5e-03  |
| 71  | Angiopoietin  | Group |           | 9.6e-03  |
| 72  | ALDH1       | Transporter |           | 9.7e-03  |
| 73  | IL13        | Cytokine  |           | 9.8e-03  |
| 74  | TNFR2       | Transporter |           | 9.9e-03  |
| 75  | IL12        | Cytokine  |           | 1.0e-02  |
| 76  | APOB        | Transporter |           | 1.1e-02  |
| 77  | Angiopoietin  | Group |           | 1.2e-02  |
| 78  | IL13        | Cytokine  |           | 1.3e-02  |
| 79  | TNFR2       | Transporter |           | 1.4e-02  |
| 80  | IL12        | Cytokine  |           | 1.5e-02  |
| 81  | APOB        | Transporter |           | 1.6e-02  |
| 82  | Angiopoietin  | Group |           | 1.7e-02  |
| 83  | IL13        | Cytokine  |           | 1.8e-02  |
| 84  | TNFR2       | Transporter |           | 1.9e-02  |
| 85  | IL12        | Cytokine  |           | 2.0e-02  |
| 86  | APOB        | Transporter |           | 2.1e-02  |
| 87  | Angiopoietin  | Group |           | 2.2e-02  |
| 88  | IL13        | Cytokine  |           | 2.3e-02  |
| 89  | TNFR2       | Transporter |           | 2.4e-02  |
| 90  | IL12        | Cytokine  |           | 2.5e-02  |
Figure 1: Top gene targets linked with the Lynch Syndrome was obtained from the Open Targets Platform based on the genetic association score (Darker the blue more the genetic association with Lynch Syndrome).

**Ingenuity Pathway Analysis:**
The Knowledgebase in Ingenuity Pathway Analysis (IPA) software (Qiagen, USA) was used to obtain the list of genes implicated in the LS. The canonical pathways, upstream regulators, and the differential regulation of gene networks in the LS were further deduced by applying the Fisher’s Exact Test (P<0.05) in IPA. The –log P values were plotted in the x-axis and the differentially expressed canonical pathways in the y-axis to derive top canonical pathways implicated in the LS.

**Results:**
The results of the study are summarized in table 1 and figure 1-9. The most genes associated with LS were found to be the MLH1, MSH2, MSH6, PMS2, EPCAM, TGFBR2, MLH3, FBXO11, and PRSS58. The most of the mutations were reported in MLH1 gene (n= 731, 589 pathogenic, and 142 likely pathogenic), followed by MSH2 (n=653, 546 pathogenic and 107 likely pathogenic); MSH6 (n=414, 367 pathogenic and 47 likely pathogenic); PMS2 (n=185, 144 pathogenic and 41 likely pathogenic); EPCAM (n=10, 8 pathogenic and 2 likely pathogenic); PRSS58 (n=7, 4 pathogenic and 3 likely pathogenic); MLH3 (n=3, 2 pathogenic and 1 likely pathogenic) and TGFBR2 (one pathogenic only).

**Discussion:**
Lynch syndrome is the most common hereditary CRC that account for more than 3% of all the colon cancer cases [12]. The genetic heterogeneity of this syndrome is related to the mutations in different genes especially in four mismatch repair genes; MLH1, MSH2, MSH6, and PMS2. The mismatch repair genes contribute to various cellular functions including repairing double-stranded DNA breaks, repairing or errors during DNA synthesis, anti-recombination and destabilization of DNA and apoptosis. MMR...
proteins serve the job of maintenance of genetic material therefore vital for the regulation of the cellular cycle. When the MMR protein is defective or lost altogether, it decreases apoptosis and increases cell survival. This leads to the selective growth advantage to the cells that lead to the more susceptibility to tissue-specific cancers [12].

**Figure 2:** Canonical Pathways in Lynch Syndrome. The canonical pathways, in the Lynch syndrome, were identified using the IPA. The -log P values were plotted in the x-axis and the differentially expressed canonical pathways in the y-axis to derive top canonical pathways such as mismatch repair signaling implicated in the Lynch Syndrome.
Figure 3: The pathogenic genetic variants observed in MLH1, MSH2, MSH6, PMS2, MLH3, EPCAM, and TGFBR2 genes associated with Lynch syndrome are shown in the genome browser.
The mutations in MLH1 are mostly present in the LS cases (around 50% of families), while the rest MMR genes mutations account for 40-50% of the syndrome. Different approaches have been utilized so far to detect mutations underlying this syndrome specifically in these four MMR genes (MLH1, MSH2, MSH6 and PMS2), which includes Next-Generation Sequencing based targeted sequencing, Multiplex Ligation-dependent Probe Amplification, Sanger Sequencing and array-CGH [13-15]. In our research, we have utilized Open Targets Platform (https://www.targetvalidation.org/) that allowed the prioritization of the genes based on the strength of their association with LS. The most important genes for the hereditary LS cancer were found to be the MLH1, MSH2, and MSH6. Together they account for more than 90% of the mutations in LS. MLH1 was on top with 731 mutations (589 pathogenic and 142 likely pathogenic), followed by MSH2 with 653 mutations (546 pathogenic and 107 likely pathogenic) and MSH6 with 414 mutations (367 pathogenic and 47 likely pathogenic). While the rest other genes including PMS2, MLH3, EPCAM, PRSS58 and TGFBR2 accounted for less than 10% of total mutations underlying LS. Canonical pathways in LS also confirmed MMR involvement in syndrome, while other pathways included colorectal cancer metastasis signaling and ovarian cancer signaling (Figure 2). While the analysis of upstream regulators of the target molecules yield to be the MBD4, PTTG1, CHI3L1, TP53, and MYC (Table 2 shows complete list).

Conclusion:
The current study has highlighted the mutation spectrum of different genes involved in Lynch syndrome and their association with different upstream regulators and involvement in canonical pathways. This study will further pave the way to accumulate all the data and genetic studies together for better prognostic and treatment options.

Conflict of Interest:
The authors declare no conflict of interest.

Acknowledgment:
This project was supported by the NSTIP strategic technologies program in the Kingdom of Saudi Arabia - Project No. 12-MED3078-03. The authors also acknowledge with thanks Science and Technology Unit, Deanship of Scientific Research, King Abdulaziz University for technical support.

References:
[1] Cunningham J et al. Am. J. Human. Genet. 2001 69:780. [PMID:1226064]
[2] Lynch HT et al. Arch. Intern. Med. 1966 117: 206. [PMID:5901552]
[3] Lynch HT et al. Am. J. Med. Sci. 1967 254: 322. [PMID:6054534]
[4] Aaltonen L et al. N. Engl. J. Med. 1998 338: 1481. [PMID:9593786]
[5] de la Chapelle et al. Annu. Rev. Genet. 1995 29: 329. [PMID:8825478]
[6] Ichikawa Y et al. Cancer Genet. & Cytogenet. 1999 112: 2. [PMID:10432927]
[7] Papadopoulos N. Ann. Oncol. 1999 10: 751. [PMID:10470419]
[8] Calistri D et al. Int. J. Cancer. 2000 89: 87. [PMID:10719736]
[9] Liu T et al. Genes. Chromosomes.Cancer.2000 27: 17. [PMID:10564582]
[10] Carvalho-Silva D et al. Nucleic Acids Res. 2019 47: D1056 [PMID: 30462303].
[11] Koscielny G et al. Nucleic Acids Res. 2017 45: D985 [PMID: 278996]
[12] Nallamilli BRB et al. Curr. Protoc. Hum. Genet. 2017 94: 10.12.1 [PMID: 28696559]
[13] Wu G et al. Genet. Test. 2001 5: 281. [PMID: 11960572]
[14] De Rosa M et al. Hum. Mutat.2003 21: 655. [PMID: 14961559]
[15] Hegde M et al. Genet. Med. 2014 16: 101. [PMID: 24310308]

Citation: Rasool et al. Bioinformation 15(4): 269-276 (2019)

License statement: This is an Open Access article which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited. This is distributed under the terms of the Creative Commons Attribution License

Edited by P Kangeune
