Nucleic Acid based diagnostic method development in molecular pathology

PhD Theses

Péter Becságh

Semmelweis University
PhD School of Pathology

Consultant:Dr. Erzsébet Rásó PhD,

Opponents: Dr.BotondTímár, Ph.D.
Dr.ZoltánKis, Ph.D.

President of Examination Committee: Dr. Péter Lakatos, Ph.D.
Members of Examination Committee: Dr. Attila Patócs, Ph.D.
Dr. ZoltánSzentirmay, Ph.D.

Budapest
2014

1. Introduction

Molecular pathology

Today, the molecular technology based molecular pathology is highly becoming an integral part of the spectrum of diagnostic pathology. The molecular pathology combines the classical histology and tools of molecular biology. The methods, which were to explore the mechanisms of molecular biology on the molecular level of basic research, are now implemented into the everyday practice of pathology. For doing this the prerequisite is to use the tests according the strict technical criteria for standard protocols. Because of the increased number of diagnostic, prognostic and therapeutic molecular targets the method development is an important element of the young molecular pathology.

Tumor development

In 2000, Weinberg et al summarized the main characteristics of the tumors in six points. Namely:

1. Independence from the signals of growth factors.
2. Independence from the growth inhibitory signals.
3. Endless proliferations.
4. Ability to avoid apoptosis.
5. Providing continuous angiogenesis.
6. Invasion into the tissue and the ability to metastasize.

In 2011, the authors rewrote and expanded the list to the following:

7. Bypass the effects of the immune system.
8. The presence of inflammatory response.
9. Maintenance and further develop of the genomic instability.
10. Presence of hypoxia, uncontrolled metabolism, and anaerobic environments.

These conditions are the result of complex alterations in a molecular matrix system, along variables in time and space. Due to the butterfly effect, each molecule modifications (both small and large scale) have an impact on the overall system. Because some changes in molecular level (either with physiological or pathological nature) enables completely new features for that molecule and this is the main phenomenon of the evolution of cancer. The functions are basically coded and regulated in the genome, in the genetic material of the cells. There are encoded differences, inherited mutations, variants which enhance the risk of cancer in the individual. If the molecules are involved in cell control, are not carried out their tasks properly the cell will develop some difference to prevent apoptosis and pass on new features providing the potential for independence. These events are moving towards an unstable genome organization. The process can be followed from the event of tumor formation and molecular profiling of individual stations, possible routes can be described. The sequential activation and silencing steps can be identified. These are the genomic processes, but still this is not the complete toolset. Further variations could be generated during transcription and translation as well. Those are closer to the function and require no changes in the genome. The creation of alternative pre-mRNA target cleavage fragments, of so-called "splice variants" can be a very effective way to achieve many times different functions. The original list of 10 features could be definitely expanded by the presence of the aberrant splicing as well.

2. Aim of the study

The accurate detection of nucleic acid sequence alterations has an equally important role in molecular biology research and molecular pathology diagnostic algorithm. This problem persists on transcriptomic and genomic level too. Many of these questions could be answered only by statistical evaluation of measurement, based on results obtained from a large number of samples. From the user's point of view this requires measurements those are technically easy to implement, but are still very reliable. Therefore, the development of diagnostic methodology in a significant proportion of cases is less spectacular, but this is undoubtedly a fundamental part of the molecular biology / pathology research. In my work two basic problems were investigated on DNA and mRNA level. One of which already requires solutions on daily basis, and a second, which is in the exploratory phase, but it becomes more evident that it will soon play a role in everyday medicine as well.

- There is a need to screen of DNA single base exchanges (SNPs / mutations) and small-scale deletions on genomic level in a system, with high reliability on a larger number of samples. I carried out a number of parallel testing of introduced method on a model, which has proved to be optimal. In particular, the human C-Kit gene exon 11, which is interestingly (but not uniquely) celebrates a series of the single nucleotide substitutions, longer and shorter deletions triplet deletions in a short range, which is ideal for testing the technology.
- The goal in the mRNA based work was the development of the suitable procedures for the identification of alternative splice variants. The human CD44 molecule was used as a model in which 10 variable exons persist and this fact enables the appearance of hundreds of potential isoforms. My goal was to develope a next generation sequencing technology method to help to find out if tumor type-specific CD44 splice pattern exists for melanoma and colorectal tumors, and if so, than which type of characteristic isoforms are specific for the given tumor (generation of the list of tumor-specific isoforms).

3. Materials and methods

Real-time PCR with saturating dye, HRM analysis

Methods used in this work were all PCR-based techniques. Detection of the extremely variable alterations in exon 11 of human C-Kit gene was realized by high resolution melting curve analysis (High Resolution Melting Curve Analysis, HRM).
The PCR primers used in this method amplified an approximately 135 and 171 bp region in target range, depending on the kind of sequence deviation.
For the experiments the *LightCycler*480 system was used. (Roche Diagnostics GmbH, Mannheim).

Recombinant PCR

To create artificial positive control fragments the recombinant PCR technique was used. During the amplification, long variant motifs-overlapping primers were used, followed by a separate PCR reaction amplified a constant ends in. By this method the series of variant were created for analytical testing of the probe based assays.

Probe based real-time PCR

Real-time PCR reactions were executed by three different labeling strategies. According to the four sub-regions inside the positive control, three and four probe based oligo sets were designed. The three strategies were the following: "simple probe" by three probes, *Taqman Probe* in three and four probes format, and finally, unlabeled Probes detected by HRM dye.

Real-time PCR experiments

The optimized protocol of the different detection strategies was put together according to the following scheme. To the simple probe system the *LightCycler* 480 Genotyping master was used (Roche Diagnostics GmbH, Mannheim) in the presence of asymmetric primer concentration. The unlabeled probes in combination with the *LightCycler* HRM master (Roche Diagnostics GmbH, Mannheim) used an asymmetric primer concentration, too. For the *Taqman* probes the *LightCycler* probe master (Roche Diagnostics GmbH, Mannheim) and equal primer concentrations were used. The final volume was set to 30µl, in the case of simple and unlabeled probes the level of asymmetry were 5 µMol and 10 µMol for forward and reverse primer accordingly (Metabion GmbH, Martinsried) and 5 µMol primer concentration for each in the *Taqman* probe-based reactions. The probe concentrations of simple, unlabeled and *Taqman* probes were identical, 5 µMol each. The *Taqman* system was a multicolor experiment, i.e., each of the 3 test contained the same probe concentration, but each with different labels.

In the case of simple and unlabeled probes the optimal MgCl2 concentration was 3 mMol after titration. The *Taqman* probe based system did not require the addition of magnesium separately because of the concentration provided by the mastermix. The reaction was carried out by TD PCR program on *LightCycler* 480 system (Roche Diagnostics GmbH, Mannheim).

Next generation sequencing of alternative splice variants

CD44 variants were determined by next generation sequencing technology. The measurement was carried out on Roche 454 sequencer, GS Junior Equipment (Roche Diagnostics GmbH, Mannheim). The process consists of four distinct parts, the library preparation, clonal amplification, sequencing reaction and finally the evaluation.

RNA isolation and transcription

The alternative splice variants of the human CD44 were detected from purified products of reverse transcribed and PCR amplified total RNA isolate. The process to this point has been performed by tumor progression group of SE 2 pathology.

The total RNA fraction isolated from tumor cell line indicated fresh frozen tissue samples originated from adult SCID mouse xenografts. The implanted colorectal cell lines were HT25

(M. Hendricks, Iowa), HCT116 (ICLC HTL95025) HCR31 and HT199 and the melanoma were the A2058, WM983B.

Library preparation for amplicon-based next generation sequencing

The library preparation process is designed to adjust the input DNA fragments to the chemical needs of the next-generation standard workflow. The cDNA transcribed fraction was amplified for amplicon sequencing by CD44-specific primer pairs.

The modified input nucleic acids were universally amplified clonally on the bead surface, separately in emPCR reaction. The resulting sequences were pyrosequenced in separate reaction wells inside the picotiter plate.
Products derived from cell line samples were identified by MID sequences.

Clonal sequencing on GS 454 system

The sequencing reaction takes place in a picotiter plate. The resulting camera images are automatically converted to the form of binary file with the linear sequence information, along with qualitative data. In accordance with the quality settings the reads can be stored after filtering. The reads were split according to the MIDs and converted into *FASTA* format by the system software. The FASTA files were screened and selected by specific exon- exon motifs – variant fishing. The identified exon combinations were filtered by quality, reliability and frequency. Reliability of the given reads was investigated in flowgram format.

4. Results

During the measurement of HRM, the different types of specific alterations successfully arranged into curve clusters. The resulted melting characteristics were clearly separated from the normal one.

The SNP containing sequences showed a characteristic peak in the melting curves, independently from the position, essentially preserving the original melting point value. In the cases of deletions the beginning of the dissociation is faster and thus showing higher difference. The longer the deletion the range of difference is higher.

Probe-based real-time PCR

After amplification the simple and unlabeled probe systems, the melting point analysis resulted in perfectly distinct values. Melting points –from the first up to the third probe - and lack of melting points showed the following picture:WT:67,2°C; 63,9°C; 63,3°C - DEL1:--; 63,9°C; 63,3°C - DEL2:--;--;63,3°C - DEL3:67,2°C; --; 63,3°C -DEL4:67,2°C; 63,9°C; --; DEL12: --; --; 63,3°C - DEL123: --; --; 63,3°C - DEL1234: --; --; --; - 1ACC: 65,4°C; 63,9°C; 63,3°C - 2TTG: 67,2°C; 61,4°C; 63,3°C - 3GGA: 67,2°C; 60,8°C; 63,3°C - 4CCA: 67,2°C; 63,9°C; 61,5°C - 1234ACCTTGGGACCA: 65,4°C; 58,4°C; 61,5°C.

Both methods are suitable for the detection the existence of short and longer deletions and to separate those from the wild-type sequence by melting point value shift.

The hydrolysis probe system for all regions gave yes / no answer, which could be detected by multicolor endpoint genotyping in combination of color pairs after amplification.

First the graph of the detection channels of 510 nm and 610 nm emission was tested. The DEL3, 2TTG, 3GGA pattern carrying triplets differences by strong emission on 510nm, the DEL1, 1ACC triplet differences carriers 610 nm emission, and WT, DEL4, 4CCA carrying deviations resulted high signal in both channels, which are created also a separate groups. As a function of the 510nm and 645nm's channels the DEL4, 4CCA variants separated from the zero emission at 645nm, and the DEL1, DEL2, DEL12, DEL123, 1ACC given high value at

645 nm. All of the WT, DEL3, 2TTG, 3GGA altered sequences given high signal on both 510 nm and 645 nm, so the three groups can be seen.

Identification of alternative splice variants of CD44

The identified exon combinations were presented after quality filtering. The quality of the individual sequencing signal were well investigated by comparison with the consensus read, based on the flowgramm,. ***The presence of exon combinations could be identified by alignment of overlapping sequences.***

After quality filtering and selecting, the results were the following combinations of exons.
(In the CD44 cDNAsequence : exon1 = v1, exon10 = v10, exon a= beforev1 exon b = afterexon 10): exon1,exon8,exon9,exon10; exon8,exon9,exon10,exonb,exon3; exona,exon1,exon6,exon3; exona,exon1,exon4,exon5; exona,exon1,exonb; exona,exon1,exona; exona,exon1,exon6,exon7; exon3,exon3,exon4,exon5,exon6; exona,exon1,exon8,exon9; exon1,exon2,exon3; exona,exon1,exon2,exon3; exon3,exon8,exon10,exonb; exon3,exonb; exona,exon1,exon3,exon4,exon5; exon3,exona,exon1,exonb; exona,exon1,exonb,exon3; exona,exon1,exon3,exon6;exona,exon1,exon3,exon3;exona,exon1,exon3,exon4; exona,exon1,exon3,exon8; exon2,exon3,exon4,exon5,exon6;exona,exonb; exona,exon1,exon3,exonb; exon3,exona,exon1,exon3; exon3,exon6; exon3,exona,exon1,exon6; exon3,exon3; exona,exon1; exon1,exon8,exon9,exonb; exona,exon1,exon8,exonb; exon6,exon7,exon8,exon9,exonb; exon4,exon5,exon6; exon1,exon8,exon9,exon10,exonb; exon3,exon4,exon5,exon6,exon7; exon3,exon4,exon5,exon6; exon3,exon8,exon9,exonb; exona,exon1,exon10,exonb; exona; exona,exon1,exon8,exon9,exonb; exon3,exon4,exon6; exon3; exon1,exon3,exon4,exon5; exon1,exon3,exon4,exon6; exona,exon1,exon6; exon8,exon9,exon10,exonb; exona,exon1,exon3; exona,exon1,exon2; exon1,exon3,exon4,exon5,exon6; exon1,exon3; exon1,exon6, exon1,exon10,exonb; exon3,exon5,exon6; exon1,exonb; exon1,exon4,exon5,exon6; exon6,exon5,exon4,exon3,exona,exon1; exon3,exon8,exon9,exon10; exona,exon1,exon8,exon9,exon10; exon3,exon7,exon8,exon9,exon10; exon3,exon10,exonb.

The tumor-specific alternative splice variant pattern of CD44

The Tumor progression Laboratory Group of SE2nd. Department of Pathology has completed the list of CD44splice variant in genetically more different human colorectal tumor and melanoma, based on the above-described method and compared it with several other human tumors CD44 ASP (alternative splice pattern).

The results showed that a tumor type characterized alternative splicing pattern exists, which proved to be qualitatively stable during tumor progression. This CD44 pattern may also be different from the normal tissue specific pattern (which has no alternative exon expression in a significant proportion of cases).
This phenomenon seems to prove the hypothesis that the alternative splicing plays a role in the development and survival of tumors, although the way in which the pattern developed is unknown.

5. Conclusions

The chosen ***HRM real-time PCR method can detect a wide spectrum of diverse variations, so the ability to sensitively detecting the presence of variations, is given***. In terms of the resolution stated, the differences could be detected by specific signals according to their nature, ***but this method is not suitable for genotyping***.

Among the two forms of targeted probe-based detection method the ***Taqman probe based system could give yes / no answer.*** However, it ***is not capable to detect alterations beyond the exact deviations they are designed for.***

The not awaited alterations could be detected by melting peeks of hybprobe, simple and unlabeled probe based systrems. After setting a series of sequence controls, it is capable for genotyping.

The unlabeled probes are cheaper, but the level of the emission signal is weak and thus resolution of the method is lower, but for focused HRM analysis it seems suitable. ***For diagnostic purpose the Taqman and the simple probe systems are acceptable.***

The heterogeneity of sequence combinations are the main feature of RNA level differences, especially in the case of splice variants. ***The most useful method to detect these variations is the PCR targeted cDNA sequencing by clonal long read next generation system. The method cannot be used to answer quantitative questions, but for pattern identification, for profile composition, to detect the asymmetry or loss of a variant, is capable. The data of exon sequence combinations can be used in real-time PCR-based test design to make the quantitative accuracy higher.***

6. Publications

Related to the dissertation

1. Becságh P, Szakács O. SettingUp a ProbeBased, (2014) ClosedTube Real-Time PCR Assay for Focused Detection of Variable Sequence Alterations. Pathol Oncol Res. 22. [Epubahead of print] PubMed PMID: 24659043.

2. Becságh P, Varga K, Szakács O, Kopper L, Orosz Z. (2010) High resolution melting curve analysis of DNA sequence alterations of various sizes. Pathol Oncol Res.;16(3):421-6.

3. Raso-Barnett L, Banky B, Barbai T, Becsagh P, Timar J, Raso E. (2013) Demonstration of a melanoma-specific CD44 alternative splicing pattern that remains qualitatively stable, but shows quantitative changes during tumour progression. PLoSOne.;8(1):e53883.

4. Bánky B, Rásó-Barnett L, Barbai T, Tímár J, Becságh P, Rásó E. (2012) Characteristics of CD44 alternative splice pattern in the course of human colorectal adenocarcinoma progression. Mol Cancer. 14;11:83.

Not related to the dissertation

1. Orosz E, Farkas A, Ködöböcz L, Becságh P, Danka J, Kucsera I, Füleky G. (2013) Isolation of Acanthamoeba from the rhizosphere of maize and lucerneplants. ActaMicrobiolImmunolHung.;60(1):29-39.
2. Dömötör D, Becságh P, Rákhely G, Schneider G, Kovács T. (2012) Complete genomic sequence of Erwinia amylovoraphage PhiEaH2. J Virol.;86(19):10899.
3. Schmidt H, Schmidt R, Niederkorn K, Horner S, Becsagh P, Reinhart B, Schumacher M, Weinrauch V, Kostner GM. (1998) Beta-fibrinogen genepolymorphism (C148-->T) is associated with carotidatherosclerosis: results of the Austrian Stroke PreventionStudy. ArteriosclerThrombVascBiol.;18(3):487-92.
4. Deák J, Lázár A, Nagy E, Benjamin Á, Becságh P, Zalka A. (2004) Optimization of a real-time PCR method for the detection of Herpes simplex virus types 1 and 2, ACTA MICROBIOLOGICA ACADEMIAE SCIENTIARUM HUNGARICAE 51: pp. 180-181.